DOCUMENT RESUME

ED 204 329 SP 018 608

AUTHOR Hord, Shirley M., Ed.; Adams, Ronald D., Ed.
TITLE Teacher Education Program Evaluation, 1981: Theory and Practice.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
REPORT NO R&DCTE-7005
PUB DATE May 81
NOTE 119p.; Papers presented at the Annual Meeting of the American Association of Colleges for Teacher Education (Detroit, MI, February 17-20, 1981).

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Data Analysis; *Data Collection; Decision Making; *Evaluation Criteria; *Evaluation Methods; Higher Education; Institutional Characteristics; Program Costs; *Program Evaluation; Program Implementation; Scientific Methodology; *Teacher Education Programs

ABSTRACT

The first paper in this collection of symposium papers presents the results of a survey of nearly 800 educational institutions and provides information on institutional characteristics, institutions engaged in teacher education evaluation, scope of data collected, methods of data collection, relationship between size of program and method of data collection, perceived value of teacher education evaluation, perceived needs in conducting teacher education evaluation, and reported budget for evaluation. In the second paper, a discussion is given on the broad spectrum of evaluation and "don'ts" for program evaluation. The third paper offers suggestions for planning, developing, implementing, and operating an evaluation program. Ten design characteristics which should be considered for developing an evaluation program are described in the fourth paper. The fifth paper presents ways to modify costs of program evaluation and suggestions for taking advantage of existing resources. The concluding paper addresses the future of teacher education evaluation. Brief summaries report on interactions in small discussion groups at the symposium at which these papers were presented. The concluding portion of the monograph is an summary from a representative of the National Institute of Education. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED204329

TEACHER EDUCATION PROGRAM EVALUATION. 1981: THEORY AND PRACTICE

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. Walton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

editors

Shirley M. Hord
Ronald D. Adams

Teacher Education Networks

Research and Development Center for Teacher Education
The University of Texas at Austin

1981

R&DCTE

SP 018 608

The research described herein was conducted under contract with the National
Institute of Education. The opinions expressed are those of the authors
and do not necessarily reflect the position or policy of the National
Institute of Education, and no endorsement by the National
Institute of Education should be inferred.

Publication #7005 available from
the Research and Development Center for Teacher Education.

TABLE OF CONTENTS

Page

FOREWORD
Oliver H. Bown, Research and Development Center for Teacher Education, The University of Texas 1

INTRODUCTION
Shirley M. Hord, Research and Development Center for Teacher Education; Ronald D. Adams, Western Kentucky University 3

I. PRESENTED PAPERS

A SURVEY OF UNDERGRADUATE TEACHER EDUCATION EVALUATION PRACTICES, Ronald D. Adams, James R. Craig, Western Kentucky University 9

WHAT NOT TO DO IN EVALUATING TEACHER EDUCATION PROGRAMS, Lyn Gubser, National Council for Accreditation of Teacher Education 29

USE-ORIENTED EVALUATION, James R. Craig, Ronald D. Adams, Western Kentucky University 38

DESIGN CHARACTERISTICS FOR MEANINGFUL TEACHER EDUCATION EVALUATION, Jerry B. Ayers, Tennessee Technological University 49

FINANCE AND RESOURCE ALLOCATIONS RELATIVE TO TEACHER EDUCATION EVALUATION, Edell M. Hearn, Tennessee Technological University 59

WHAT IS THE FUTURE OF TEACHER EDUCATION PROGRAM EVALUATION, Gene E. Hall, Research and Development Center for Teacher Education, The University of Texas at Austin 65

II. GROUP RESPONSE STATEMENTS

TEACHER EDUCATION PROGRAM EVALUATION, A SEARCH FOR ALTERNATIVES, Tom Savage, Texas A&M University 79

PROGRAM EVALUATION: WHAT, WHEN, AND HOW?, Ted White, Glassboro State College 82

PROGRESS IN EVALUATION OF TEACHER EDUCATION, Vernon Dravland, University of Lethbridge 85

Page

GROPING... AND COPING... WITH CONTRADICTORY PROGRAM EVALUATION REQUIREMENTS, Gary G. deVoss, The Ohio State University 88

FACULTY REACTIONS TO PROGRAM EVALUATION: A LITANY OF "YES, BUT'S..." Len Froyen, University of Northern Iowa 91

THE USE OF PROGRAM EVALUATION RESULTS: PROBLEMS AND SUGGESTIONS, Ann Hungerman, University of Michigan . . . 94

PROGRAM EVALUATION--A SYSTEMS APPROACH, James J. Garland, Southwest Texas State University 97

SMALL COLLEGE NEEDS IN FOLLOW-UP EVALUATION STUDIES, Al Kilgore, University of Nebraska-Lincoln 100

III. SUMMARY

So WHAT IS EVERYBODY TELLING US ABOUT EVALUATING TEACHER EDUCATION? Joe Vaughan, National Institute of Education 103

FOREWORD

The R&D Center's mission is to design and conduct research and related activities in the field of teacher education. This is carried out in close collaboration with practitioners, policy makers and other researchers who are most cognizant of emergent unsolved problems and challenges in the field. Through its own research, the Center is contributing to the knowledge bases underlying effective teaching and learning in various college, school and classroom contexts, effective teacher education systems throughout the career cycle, and successful implementation of research-based organizational and instructional practices.

In keeping with its role as one of the national centers for research and development in education, the Center attempts to facilitate and coordinate various planning, networking and collaborative activities which link communities of scholars and practitioners both nationally and internationally. One interesting example of an informal and spontaneous network which has developed over the past few years has involved individuals from across the country working in relative isolation on the design and conduct of teacher education program evaluation follow-up studies. This network (TEPFU) was officially "born" in April, 1978, when the Center hosted a national conference of individuals who had conducted such programs or studies. Conference proceedings were published, and the Center began to receive numerous inquiries for further information and requests to be affiliated with further meetings and other information sharing opportunities. A newsletter was developed as one way of responding to this increasing interest. Over the past three years, meetings of the expanding group have been held in conjunction with annual meetings of appropriate scholarly and professional associations, and these have provided research reports, professional dialogue and increased membership in the network.

We believe that increased attention to systematic evaluation of teacher preparation programs including appraisal and analysis of graduates' performance and experience through their first professional year is an invaluable source of feedback in the redesigning, refinement and enrichment of preparation programs. We are noting increased sophistication in the design of such studies and in solving the many logistical and economic problems which attend the implementation effort. The R&D Center is pleased to play its part in facilitating communication and some emerging synergy in the community of scholars engaged in this important work.

Oliver H. Bown
Director
Research and Development Center for Teacher Education
The University of Texas at Austin

INTRODUCTION

An increasingly important problem faced by schools, colleges and departments of education is evaluation of their teacher preparation programs. Internal pressures encourage institutions to demonstrate program effectiveness; external mandates for meeting accreditation standards reinforce the need to engage in evaluation. Without well developed program assessment followed by systematic implementation of the changes warranted by the examination, it is difficult to imagine how program effectiveness may be improved.

Over the last decade there has been a slow but growing interest and involvement in teacher education program evaluation. The Sandefur (1970) model provided a cornerstone for the development of evaluation studies early in the 1970's by institutions such as Western Kentucky University (1972) and Tennessee Technological University (1974). In several other areas of the nation and in Canada (Dravland & associates), others were also beginning to develop strategies for assessing teacher education programs. The National Council for Accreditation of Teacher Education (NCATE) standards further encouraged--insisted on--program examination.

Then in 1978 more focused attention occurred through various organized structures. The annual meeting of the American Association of Colleges for Teacher Education (AACTE) that year devoted one of eight program "strands" which cut across the four day meeting to follow-up studies of teacher preparation programs. In the spring of that year, the Texas Research and Development Center for Teacher Education became involved in program evaluation efforts in a response to the needs of individuals and institutions engaged in evaluation and follow-up studies. A collegium of eight institutions met under the auspices of the Texas R&D Center in order to share techniques, concerns, problems and successes. This small but very successful attempt at linking persons involved in program evaluation provided the forum for institutions to share what they were doing. Several results of this working conference contributed to the increased viability of program evaluation.

A monograph containing the reports of the eight institutions' evaluation studies was published (Hord & Hall, 1979). This volume has had wide dissemination from the R&D Center and provides a practical reference for how-to-do such studies, as reported by those who have had extensive experience. A second important outcome of the conference was the initiation of a network for persons who wished to communicate and continue to share with each other. This network is known as TEPFU, Teacher Education Program Follow-Up. A very informal newsletter, published three times a year by the R&D Center, was instituted, and serves as a means for sharing information and disseminating products to the TEPFU network members. The membership of this network continues to expand.

Following these activities in 1978, there was an interest in, but a lack of funds for supporting ways that would permit individuals to engage in personal interaction and dialogue. While the newsletter kept people connected, there was lacking a vehicle for making discussion possible. Consequently, small ad hoc gatherings of program evaluators (and those interested in so becoming) were held in conjunction with various professional organization meetings.

In 1980, external pressure on institutions to do program evaluation was escalating. At the annual AACTE meeting that year the small space accorded a session on follow-up evaluation of graduates, anchored by Adams and Ayers, could not accommodate the large number of people trying to attend. Persons who could get in indicated interest in staying in touch and participating in some inexpensive semi-formal strategies for the purpose of sharing with each other and acquiring information related to program evaluation. Several months later at the annual American Educational Research Association (AERA) meeting, a group of persons again met informally to generate such ideas; some very promising strategies for making program evaluation efforts more visible were identified and some activities to respond to the needs and interests of persons involved in or expecting to be involved in program evaluation were suggested. These events reflected a substantial vigor in this area of inquiry.

Because of keen interest and responses from the teacher education program evaluation network that reflected continuing need for assistance, a session was designed to be a part of the 1981 AACTE annual meeting. As a result of the 1980 experience a better case was made for more space and time to be available for TEPFU activities. Thus, a program format that included a large block of time was arranged in order to address issues (e.g., accreditation), share experience and expertise, and engage participants in meaningful activities targeted at improving their own skills and capabilities for undertaking their program assessment. The AACTE staff and program committee facilitated the arrangement of program time and meeting space for such a session.

The session was designed in two parts: (1) six invited papers focusing on issues of program evaluation were presented by the authors; (2) all session attendees were organized into small discussion groups for interaction and dialogue, stimulated by the paper presentations. Members of the TEPFU network guided small group discussions. As an outcome of each small discussion group, a statement of response reflecting their insights, reactions, and questions was to be prepared. These statements collaboratively produced by participants of both large and small institutions, by persons with some years of experience and others with none, would represent current experience and thinking "in the field."

The publication of this monograph, Teacher Education Program Evaluation, 1981: Theory and Practice, further testifies to the importance of that session. It includes the six formally prepared papers and the summary papers which resulted from the work of the small interactive groups. This small volume will acquaint the reader with current key issues in teacher education program evaluation as well as provide assistance for those engaged in planning, implementing or operating evaluation programs.

Part I

Part I of this monograph is the set of papers prepared for presentation at the session. The first paper by Adams and Craig reports on current practice in evaluation efforts. A survey was developed at Western Kentucky University and sent to nearly 800 institutions in order to reveal the state of the practice in program evaluation. Responses to the survey are reported by the authors in eight categories: institution demographics, institutions engaged in teacher education evaluation, area of data collected, methods of data collection, relationship between size of program and method of data collection, perceived value of teacher education evaluation, perceived needs in conducting teacher education evaluation and reported budget for evaluation.

In the second paper, Lyn Gubser suggests "a dozen don'ts" for what not to do in program evaluation. From his perspective as director of the National Council for Accreditation of Teacher Education, Gubser discusses the broad spectrum of evaluation and summarizes with how-not-to-do-its for people interested in establishing successful program evaluation.

In the third paper, Craig and Adams suggest a process for planning, developing, implementing, and operating an evaluation program that will "maximize the utilization" of evaluation data in program decision making.

A critical aspect of the design suggested by the Craig and Adams paper is more specifically addressed in the fourth paper by Jerry Ayers, Tennessee Technological University. Ten design characteristics which should be considered for developing an evaluation program are described, as well as problems which may be encountered if the characteristics are left unregarded.

Edell Hearn reports on Tennessee Technological University's evaluation system and its costs. Effective evaluation of programs should allow appropriate modifications in programs, but this can happen only if finance and resource allocations for the evaluation are part of the institution's budget. Ways to modify costs and suggestions for taking advantage of existing resources are presented.

The concluding paper addresses the future of teacher education evaluation. Gene Hall, Research and Development Center for Teacher Education, raises questions about the future of teacher education program evaluation and follow-up studies as a whole. He distinguishes between issues that are controllable and those that are not. He makes concrete recommendations for the future regarding the controllable areas, and speculates on the pressures which may be anticipated in the areas not under control.

Part II

Included in this section of the monograph are the brief summary reports of the interaction by the participants in the small discussion groups. Their reactions to issues raised by the paper presenters were synthesized by the group facilitator. More than 200 persons attended the session and 70 remained into the dinner hour to participate in the group activity.

Part III

The concluding portion of this monograph is an executive summary by Joe Vaughan of the National Institute of Education. Vaughan provides a synthesis of themes of the papers and group statements to establish a composite image of the issues, questions and concerns which emerge from all the writings.

It is the hope of all who were involved in contributing to this monograph that it will prove to be instructive, providing useful information to program evaluation practitioners. Each of the authors invites the readers to contact them for further dialogue and interaction. Despite the difficulties of doing program evaluation, it is apparent that for the present, teacher education program evaluation lives.

Shirley M. Hord
Research and Development Center for Teacher Education

Ronald D. Adams
Western Kentucky University

May, 1981

A SURVEY OF UNDERGRADUATE TEACHER EDUCATION EVALUATION PRACTICES

Ronald D. Adams
James R. Craig
Western Kentucky University

I. INTRODUCTION

Several factors have led to our conduct of a survey of teacher education evaluation practice. First, due to our experience in teacher education evaluation since the Fall of 1971, we have had the opportunity for dialogue with numerous individuals who are planning or are engaged in evaluation programs. Many of these individuals have expressed a desire to know the state-of-the-practice in teacher education evaluation.

Second, the growth of the TEPFU (Teacher Education Program Follow-Up) group and the high attendance at national meetings where teacher education evaluation topics were presented has demonstrated a continuing interest in this topic. One of the recommendations from the TEPFU group was that a study be conducted to determine the current evaluation practice in teacher education evaluation as there appears to be little research in the professional literature on this topic.

Third, the NCATE standards for accreditation have explicitly stated that institutions producing teachers must have a systematic evaluation of their teacher education programs to include gathering data while students are in the preservice program and after they have entered the teaching profession. Evidence must also be provided of utilization of data for program improvement. Undoubtedly, these standards have provided much of the interest in teacher education evaluation.

Finally, in a 1978 doctoral dissertation at the University of Tulsa, Wennette Pegues reported a survey of teacher education follow-up evaluation practice. This study provided a beginning for determining the status of teacher education evaluation practice. Two of her findings have particular interest to this study. First, she concluded that, "There is only moderate comtemporary involvement in systematic follow-up evaluative practices by teacher education programs." Only half of the respondents reported that they were engaged in this practice. Second, she concluded that, "There is a minimal level of evaluative sophistication in the design and the implementation of comtemporary evaluative schema." The most prevalent schema was the controversial one-shot questionnaire sent to graduates the first year following graduation.

From the preceding discussion, it seemed clear that a current survey of teacher education evaluation practice could provide meaningful information in identifying needs and future directions in this area. An initial survey form was developed by the authors to obtain information from the following areas:

1) Institutional demographics
2) Topics on areas evaluated
3) Techniques used in evaluation system
4) Sources of evaluation data
5) Points of evaluation
6) Needs relative to evaluation practice
7) Value placed on evaluation
8) Resources devoted to evaluation

The initial survey form was reviewed by college administrators at WKU and a later revised version was sent to selected members of the TEPFU group, via assistance from the R&D Center for Teacher Education. Finally, the third version was sent to AACTE and NCATE officials for input.

All suggestions were considered, final changes were made, and the survey forms were mailed about the middle of November, 1980. Even with a short return time, the holiday season, and the rather lengthy form, a 36 percent return rate was obtained by the end of December. A follow-up survey was sent about the second week in January, 1981, resulting in an additional 20 percent return for a total of 442 returned survey forms to date.

To prepare for this report, data obtained from 397 forms (51 percent) were processed. The remaining will be reanalyzed for a later publication. This more complete report will contain detailed analyses and findings and should be available during the summer of 1981.

Time constraints do not permit presenting in great detail the results of this survey--in fact, much of the data have yet to be analyzed beyond the rudimentary findings prepared for this paper. However, the findings presented here should provide a basis for understanding more fully the present state-of-the-practice in teacher education evaluation, and suggest directions we must take as a profession if we are to improve this essential practice.

II. FINDINGS

The findings presented for this report represent preliminary analyses of the data obtained from 397 responding institutions. Thirty-eight additional questionnaires have been received since these analyses were conducted and are being prepared for inclusion in the data set. Nine institutions did not qualify for inclusion for analysis due to such factors as having no undergraduate programs or no longer being in operation. The accounted for responses totaled 444 for a return rate of 57.0 percent.

The findings are presented in eight categories:

A. Institution Demographics
B. Institutions Engaged in Teacher Education Evaluation
C. Area of Data Collected
D. Methods of Data Collection
E. Relationship Between Size of Program and Method of Data Collection
F. Perceived Value of Teacher Education Evaluation
G. Perceived Needs in Conducting Teacher Education Evaluation
H. Reported Budget for Evaluation

Undoubtedly, additional analyses will provide more information regarding the state-of-the-practice in teacher education evaluation than these preliminary analyses. However, we feel findings from these analyses provide more insight into what is being done in teacher education evaluation than has been known to date.*

---

*Due to space limitations the tabular presentation of data will be limited in this report.

A. Institution Demographics

The responding institutions were evenly split between public and private with 52.4 percent public and 46.3 percent private. NCATE accredited institutions comprised 63.7 percent of the sample while 82.6 percent reported membership in AACTE. Geographic region represented by responding institutions was indicated by membership in regional accrediting agencies and can be seen in Table 1.* Institutions in the North Central Association comprised 48.4 percent of the respondents, followed by the Southern Association with 27.4 percent and the Middle States Association with 11.5 percent. These data are very close to the percentages obtained by Pegues in 1978.

Institutional size ranged from 300 to 63,115, with undergraduate teacher education programs ranging from 13 to 7,300 students. These data are presented in Table 2. It can be observed that 57.7 percent of the programs have 500 or less enrollment. Institutions producing only Bachelor's Degrees made up 31.8 percent of the sample while 22.0 percent of the responding institutions offer the Doctorate.

B. Institutions Engaged in Teacher Education Evaluation

Two categories of data, size of undergraduate teacher education program and public/private classification were cross tabulated with institutions' response to Question 1, Part II of the survey form. This question asked institutions to indicate if they were engaged in evaluation of their undergraduate teacher education programs. From the responses it was determined that 88.7 percent of the institutions indicated some evaluation activity.

A cross tabulation of size of program and public or private classification indicated a relatively strong relationship between these two variables. Smaller programs tend to be in private institutions while larger programs tend to be in public institutions. While this relationship is not surprising, it should be kept in mind when discussing other analyses.

Three categories of size were utilized in the cross tabulation analysis with those reporting evaluation practices. They were as follows:

small - 1 to 250 students
medium - 251 to 1,000 students
large - 1,001 to 7,500 students

---

*Figures appear at the end of each paper.

There appeared to be little difference in the percentage of institutions reporting evaluation practice and the size of the teacher education program.

Similarly, when public or private classification was cross tabulated with reported evaluation practice no meaningful differences were noted. This was expected due to the relationship between program size and public/private classification.

C. Area of Data Collection

Institutions were asked to respond to a list of seven evaluation areas for which they collected data. They were instructed to check each area for which they collected data. The "Teaching Skills, Strategies, and Techniques" received the most checks, with "Knowledge of Subject Matter" and "Relationships with Students" following, respectively. The least checked items were "Relationship with Parents" and "Participation in Professional Organizations", respectively.

Additional analyses are planned to determine if there are relationships between evaluation areas and other variables. For example, the question "are method and area related?" may provide some insight into source of data collection and what is being collected.

D. Methods of Data Collection

The major focus of the survey was to determine the state-of-the-practice in teacher education evaluation. A major component was the methods employed to collect data to include type of data collecting devices or procedures, source of data and frequency or points of data collection. Institutions were asked to respond to seven data collection procedures each having a matrix of data source by collection frequency. Through this reporting procedure it could be determined the most often utilized source and the frequency with which data were collected from sources for any given procedure.

The seven data collection procedures contained in the survey were:

1. Questionnaires
2. Personal Interviews
3. Direct Classroom Observation
4. Media-Aided Work Sample Observation
5. Standardized Tests
6. Professional Competence Measures
7. Personal Characteristics and Attitudes Measures

Respondents were asked to place the approximate percent of program enrollment sample for each data collection period and each source. While the majority of respondents followed these directions, several returned surveys had only checks in the matrix grids. Thus, for these preliminary analyses, sample size was not considered; rather, the frequency of participation only was established.

Data from these questions provided "maps" of evaluation practice that describe the methods utilized in teacher education evaluation. While these analyses provided only summary data on methods of evaluation, they allow for some generalizations to be drawn.

Questionnaires were by far the most utilized form of collecting evaluation data, both in preservice and follow-up. The source most often given was the student/graduate and the most frequent point of data collection was at exit from program (65.8 percent). The supervisor was the second most used source for evaluation data with peers providing some data in the preservice program. Table 3 presents these data.

When a usage criterion of 10 percent was utilized to examine the results, it was determined that data were collected from student/graduate and supervisor at entry and continued through the 4-6 year of follow-up. Peer data were only used in the preservice program in categories "during" and "exit". The heavy black line indicates the cells which met the usage criterion.

Personal interviews were used to a much lesser extent than questionnaires, and were limited primarily to use in the preservice program. The student/graduate and supervisor were again the primary sources. Table 4 contains the summary data on personal interview.

Use of direct classroom observation as a technique for collecting evaluation data was quite limited except in preservice programs. Students/ graduates were observed in many programs during and at the end of teacher education programs, and to a very limited extent during the first year of teaching. Pupils were observed mostly during the undergraduate preparation program. These data are presented in Table 5.

The method using media-aided work sample observations produced a 10 percent usage in only the "during" category of the preservice program. Table 6 shows these data.

Use of standardized tests were also quite limited and were restricted to use in the preservice program. This type of data was primarily collected at entry and at exit to the undergraduate teacher education program, as can be seen in Table 7.

The last two categories of data collection provided quite similar results. Professional competence and personal characteristics and attitudes were probably recorded as extensions of questionnaires by many respondents. They reflect usage during the preservice and first year follow-up with limited usage the third and fourth-sixth years of follow-up. Primary sources were the student/graduate and supervisor. Limited usage of peers as a source was noted during preservice. Tables 8 and 9 contain the summary data for these variables.

In summary, all techniques were used to varying degrees in preservice evaluations. The sources, however, were limited to the student/graduate, supervisor, and to a lesser degree peers. Follow-up evaluation was conducted less often than preservice evaluation. Also, fewer techniques were widely used for follow-up evaluation with questionnaires being by far the most often reported technique. As with preservice, the sources were mainly limited to student/graduate and supervisor.

E. Relationship Between Size of Program and Method of Data Collection

Given the extreme range in size of teacher education programs reported, the authors decided to investigate the relationship between size of program and methods of evaluation reported. Chi square analyses were utilized to determine if a relationship existed between the two sets of variables.

First, cross tabulations were made between size (broken into three categories) and overall indication of use for a given method or technique. Chi squares were computed for each of the seven techniques and size. Only two of these analyses approached significance at the .05 level of confidence. It was observed that direct classroom observation and media-aided work sample observations obtained a large enough chi-square value to approach or exceed the significance level of .05 when compared to program size. It would appear that in both cases, the smaller programs reported more use of these techniques than the larger programs.

A second set of analyses dealt only with follow-up data. Due to the small frequency of use, only questionnaires, personal interviews, and direct classroom observations were included in these analyses. However, the analysis covered the first, third and fourth-sixth years of follow-up for each technique.

From these analyses an interesting finding emerged. There appeared to be a pattern of usage for techniques and program size. Four of the five chi square analyses were significant beyond the .05 level of confidence for the first year follow-up data. There appeared to be more usage among the larger programs of questionnaire, personal interview, and direct classroom observation than among the smaller programs. This was an opposite indication as found for direct classroom observation and media-aided work sample observation for total evaluation usage.

It also appeared that by the third year and beyond, the usage of these techniques diminished to the point of no differences between programs of varying size, i.e., for each size of program classification, there was little being done for follow-up evaluation.

F. Perceived Value of Teacher Education Evaluation

AACTE institutional representatives or their designees were asked to complete the survey form. It was noted that for the most part administrators completed the questionnaire. In the "other" category, the most often given titles of the respondents were director of student teaching or director of the research unit. Thus, the perceptions of the value of teacher education evaluation was primarily from an administrative viewpoint.

Perceptions of the value of teacher education evaluation were compared for those institutions who had evaluation and those who reported no evaluation activities. While no formal statistical analysis was computed to determine the significance of the difference, it appeared that the institutions having evaluation programs valued teacher education evaluation more than did the non-evaluation institutions. However, both groups tended to perceive evaluation as more positive than negative.

Additional analyses were conducted to determine if size of program was related to perceived value of evaluation. Chi square analysis was utilized with both the institutions reporting evaluation programs and those not reporting evaluation programs.

In both cases there appeared to be no relationship found for administrators' perception of value of evaluation and size of program. However, in both groups there was a significant chi square found for perceived value of evaluation by faculty and program size. Faculty in the smaller programs were rated as valuing evaluation more than faculty in the larger institutions for both groups. Students were also rated as valuing evaluation more for those institutions which reported having evaluation programs. It should be kept in mind, however, that the raters for the most part were administrators.

G. Perceived Needs in Conducting Teacher Education Evaluation

Respondents were also asked for perceived needs in conducting evaluation of teacher education programs. Again, those responding as having evaluation programs and not having evaluation programs were informally compared. "Developing Instrumentation" was a frequently given need for both groups, ranked 3rd for the evaluation group and 1st for the non-evaluation group.

Assistance with "Planning and Implementing a Feedback System" and "Utilizing Data for Program Improvement" appeared to be the greatest need for institutions having evaluation programs, while "Planning an Evaluation System" and "Identifying Instrumentation" ranked number two for institutions not having an evaluation system.

H. Reported Budget for Evaluation

To estimate the resources being allocated to teacher education evaluation, institutions were asked to give the approximate operating budget for evaluation of teacher education programs. The range in budget allocations was from $15 to $81,610. Slightly more than 50 percent of the institutions reported spending less than $1,250 on evaluation. However, 18.5 percent reported spending over $10,000. The mean and median for all institutions were $6,280.20 and $1,202.50, respectively.

Institutions were also asked to indicate the source of funding for their evaluation program. Of those responding, 9.4 percent or 32 institutions indicated some or total external funding from public sources. In addition, 1.2 percent or four institutions reported funding from private funds. Most of the institutions that reported funding source (n=239), indicated total internal funding for their evaluation programs (92.9 percent). Those not reporting funding source totaled 100. Of those institutions reporting funding sources, 23.3 percent indicated funding on a one year basis, while 71.1 percent indicated funding was recurrent.

III. SUMMARY FINDINGS

The following conclusions were made from the preceding preliminary analysis:

1. There appeared to be little difference in the percentage of institutions reporting evaluation practice and the size of enrollment in the teacher education program or whether the institution was public or private.

2. The areas of "Teaching Skills, Strategies, and Techniques", "Knowledge of Subject Matter" and "Relationships with Students" were the most frequently evaluated.

3. Questionnaires were by far the most used method of collecting data with the student/graduate the most frequent reported source of data.

4. More evaluation practice and more varied methods of evaluation were reported for preservice evaluation than for follow-up evaluation.

5. Smaller institutions tended to use Direct Classroom Observations and Media Aided Work Sample Observations more than larger institutions. However, this technique was most often reported in preservice evaluation practice.

6. When only follow-up evaluation was considered, there were differences noted the first year, but not the third or fourth-sixth years for Questionnaires, Personal Interview, and Direct Classroom Observation. The larger institutions appeared to be engaged more often than the smaller institutions for the first year follow-up, but no differences emerged after the first year. There was a tendency for smaller institutions to do more follow-up at the fourth-sixth year point.

7. From the college administrator perspective, the following values were observed:

    a. Institutions who responded that they have evaluation programs, valued evaluation practice more highly than institutions who reported no evaluation programs.

   b. Size of program was not related to the degree administrators valued evaluation practice.

   c. Faculty and students in smaller institutions were perceived to value evaluation practice higher than their counterparts in larger institutions.

8. Programs that reported evaluation practice most often perceived the following needs, respectively:

   a. Planning and Implementing a Feedback System.

   b. Utilizing Data for Program Improvement.

   c. Developing Instrumentation.

9. Programs that did not report an evaluation program perceived the following needs most frequently:

   a. Developing Instrumentation.

   b. Identifying Instrumentation.

   c. Planning an Evaluation System.

10. Programs reporting operation budgets varied greatly in the amounts of support. In nearly all cases funding was from internal funds and most reported recurring funding.

TABLE 1

Summary of Institutions by Selected Demographic Categories

Category	f	%
Public	208	52.4
Private	184	46.3
NCATE accredited	253	63.7
AACTE member	328	82.6
Regional Accrediting Assoc.		
New England	17	4.4
Middle States	45	11.5
North Central	192	48.4
Northwest	20	5.1
Southern	107	27.4
Western	9	2.3

TABLE 2

Summary of Institutions Responding by Size of Enrollment

Enrollment	300 to 1,300	1,400 to 3,600	3,779 to 6,500	6,808 to 10,000	10,026 to 20,000	20,142 to 63,115
Total Inst.						
%	25.6	24.4	16.6	11.1	15.0	7.3
Cum %	25.6	50.0	66.6	77.7	92.7	100.0
Under Grad. Teacher Ed. Program	13 to 100	104 to 250	255 to 500	525 to 1,000	1,100 to 2,000	2,072 to 7,300
%	14.0	21.1	22.6	19.7	15.8	6.8
Cum %	14.0	35.1	57.7	77.4	93.2	100.0

Valid cases = 385

TABLE 3

Summary Analysis of Institutions Utilizing _Questionnaires_ in Evaluation of Teacher Education Programs

Questionnaires

		Preservice Program			Follow Up Year					
Data Source		Entry	During	Exit	1	2	3	4-6	7-10	11+
The student/ graduate	f	119	169	223	180	71	96	96	12	5
	%	35.1	49.9	65.8	53.1	20.9	28.3	28.3	3.5	1.5
A Supervisor (advisor, principal, dept. supervisor, etc.)	f	57	145	153	151	51	71	66	7	4
	%	16.8	42.8	45.1	44.5	15	20.9	19.5	2.1	1.2
Peer Teachers	f	10	53	38	11	3	7	7	0	0
	%	2.9	15.6	11.2	3.2	.9	2.1	2.1	0	0
Pupils	f	4	26	13	8	4	6	5	0	1
	%	1.2	7.7	3.8	2.4	1.2	1.8	1.5	0	.3
Parents	f	2	5	3	0	0	0	0	0	0
	%	.6	1.5	.9	0	0	0	0	0	0
Other	f	7	9	12	7	1	2	0	0	0
	%	2.1	2.7	3.5	2.1	.3	.6	0	0	0

TABLE 4

Summary Analysis of Institutions Utilizing Personal Interviews in Evaluation of Teacher Education Programs

Personal Interviews

		Preservice Program			Follow Up Year					
Data Source		Entry	During	Exit	1	2	3	4-6	7-10	11+
The student/ graduate	f	88	105	87	32	10	9	17	2	3
	%	26	31	25.7	9.4	2.9	2.7	5.0	.6	.9
A Supervisor (advisor, principal, dept. supervisor, etc.)	f	37	91	84	35	12	12	16	2	3
	%	10.9	26.8	24.8	10.3	3.5	3.5	4.7	.6	.9
Peer Teachers	f	6	35	26	6	1	2	1	1	0
	%	1.7	10.3	7.7	1.7	.3	.6	.3	.3	0
Pupils	f	3	10	5	2	2	1	0	0	0
	%	.9	2.9	1.5	.6	.6	.3	0	0	0
Parents	f	1	3	3	2	2	1	0	0	0
	%	.3	.9	.9	.6	.6	.3	0	0	0
Other	f	2	6	3	1	0	0	0	0	0
	%	.6	1.7	.9	.3	0	0	0	0	0

TABLE 5

Summary Analysis of Institutions Utilizing Direct Classroom Observations in Evaluation of Teacher Education Programs

Direct Classroom Observations

Data Source		Preservice Program			Follow Up Year					
		Entry	During	Exit	1	2	3	4-6	7-10	11+
The student/ graduate	f	55	225	139	34	12	11	11	3	2
	%	16.2	66.4	41.0	10.0	3.5	3.2	3.2	.9	.6
Pupils	f	9	42	29	12	6	5	4	1	1
	%	2.7	12.4	8.6	3.5	1.8	1.5	1.2	.3	.3
Other	f	7	28	19	5	1	0	2	0	0
	%	2.1	8.3	5.6	1.5	.3	0	.6	0	0

TABLE 6

Summary Analysis of Institutions Utilizing
Media-Aided Work Sample Observations in Evaluation
of Teacher Education Programs

Media-Aided Work Sample Observations

		Preservice Program			Follow Up Year					
Data Source		Entry	During	Exit	1	2	3	4-6	7-10	11+
The student/ graduate	f	16	158	33	5	1	3	0	0	1
	%	4.7	46.6	9.7	1.5	.3	.9	0	0	.3
Pupils	f	3	22	9	2	0	1	0	0	0
	%	.9	6.5	2.7	.6	0	.3	0	0	0
Other	f	0	8	1	0	0	0	0	0	0
	%	0	2.4	.3	0	0	0	0	0	0

TABLE 7

Summary Analysis of Institutions Utilizing Standardized Tests in Evaluation of Teacher Education Programs

Standardized Tests

		Preservice Program			Follow Up Year					
Data Source		Entry	During	Exit	1	2	3	4-6	7-10	11+
The student/ graduate	f	46	28	99	8	5	6	3	0	0
	%	13.6	8.3	29.2	2.4	1.5	1.8	.9	0	0
Pupils	f	0	2	3	0	0	0	0	0	0
	%	0	.6	.9	0	0	0	0	0	0
Other	f	0	0	0	3	0	0	0	0	0
	%	0	0	0	.9	0	0	0	0	0

TABLE 8

Summary Analysis of Institutions Utilizing Professional Competence in Evaluation of Teacher Education Programs

Professional Competence

		Preservice Program			Follow Up Year					
Data Source		Entry	During	Exit	1	2	3	4-6	7-10	11+
The student/ graduate	f	47	138	154	95	30	43	39	3	2
	%	13.9	40.7	45.5	28.0	8.8	12.7	11.5	.9	.6
A Supervisor (advisor, principal, dept. supervisor, etc.)	f	38	151	147	107	29	45	43	3	1
	%	11.2	44.5	43.4	31.6	8.6	13.3	12.7	.9	.3
Peer Teachers	f	12	42	34	13	2	10	4	0	0
	%	3.5	12.4	10.0	3.8	.6	2.9	1.2	0	0
Pupils	f	5	15	15	10	4	8	5	0	0
	%	1.5	4.4	4.4	2.9	1.2	2.4	1.5	0	0
Parents	f	2	4	2	1	1	1	0	0	0
	%	.6	1.2	.6	.3	.3	.3	0	0	0
Other	f	2	9	9	2	0	1	0	0	0
	%	.6	2.7	2.7	.6	0	.3	0	0	0

TABLE 9

Summary Analysis of Institutions Utilizing Personal Characteristics and Attitudes in Evaluation of Teacher Education Programs

Personal Characteristics

Data Source		Preservice Program			Follow Up Year					
		Entry	During	Exit	1	2	3	4-6	7-10	11+
The student/ graduate	f	71	126	129	65	17	33	28	4	2
	%	20.9	37.2	38.1	19.2	5.0	19.7	8.3	1.2	.6
A Supervisor (advisor, principal, dept. supervisor, etc.)	f	53	127	115	75	17	34	33	3	1
	%	15.6	37.5	33.9	22.1	5.0	10.0	9.7	.9	.3
Peer Teachers	f	13	41	30	9	2	4	1	0	0
	%	3.8	12.1	8.8	2.7	.6	1.2	.3	0	0
Pupils	f	3	9	12	8	3	6	4	0	0
	%	.9	2.7	3.5	2.4	.9	1.8	1.2	0	0
Parents	f	2	4	2	1	1	1	0	0	0
	%	.6	1.2	.6	.3	.3	.3	0	0	0
Other	f	7	11	6	2	0	0	0	0	0
	%	2.1	3.2	1.8	.6	0	0	0	0	0

WHAT NOT TO DO IN EVALUATING TEACHER EDUCATION PROGAMS

A DOZEN DON'TS

Lyn Gubser
National Council for Accreditation of Teacher Education

Being someone who tries to take a positive approach to things and who enjoys a natural tendency to rebel against lists of "don'ts," I was a bit put off by the request that I begin our discussion of teacher education evaluation by describing what *not* to do. I find it saddening to study program evaluation efforts that have failed to fly--crashing and burning after takeoff, if they even got off the ground. But if our task is to examine the wreckage of failed efforts, then at least let us conduct our investigation with an eye out for clues to future success.

The Lay of the Land

Evaluation of teacher education involves the appraisal of the entirety of any system that prepares school personnel. It includes the assessment of programs that prepare counselors, school psychologists, administrators, media technologists, librarians, and assorted specialists, in addition to instructional personnel. In planning for the evaluation of professional programs, one should not overlook the extensively broad range of specialties that professional education has come to include.

If we regard programs as systems, then we must be concerned with the evaluation of inputs, processes, and outputs. Attention should be paid to the evaluation of such factors as program admission and retention, of clinical facilities and resources, of faculty competence and experience, in addition to the assessment of qualities of those who complete professional programs. One of the most common malpractices in teacher education evaluation, as we shall see, is the frequent confinement of assessment efforts to the evaluation of graduates.

Sound program evaluations can be constructed on any number of philosophical and contextual foundations. Program review is one approach and is somewhat analogous to calling together a group of chefs to evaluate the baking of a cake. From their combined knowledge and experience decisions are made on the amount of flour, sugar, and flavorings to add to the recipe to obtain the desired result. But decisions must also be made on process variables, such as how hot one should keep the oven. Quality control is a continuous procedure in baking. Although the ultimate question is how tasty is the cake, a multitude of evaluations are made before that question can be asked. Even then there remain such long-range questions as how long can the cake be kept, under what conditions, and how best should it be served?

We in education have tended to focus on what goes into the cake. All of us have had our pet suggestions, most of which have been unsupported by reliable data. Only within the past decade have we actually tried to "taste the cake." For many reasons our successes have been infrequent.

Educators have only rarely sought to assess in-put variables or evaluate processes in empirical ways. Few evaluation models have attempted to systematically assess textbooks or establish a sequence of program content through careful investigation. We are hard-pressed to comment upon the successes or failures of these evaluation approaches, for one cannot effectively criticize that which only rarely exists. If enrollments are up and the central administration is happy, that's been good enough for many of us, and we have pushed student teachers out of the academic nest and into the classroom. If they haven't returned, we assume that they flew away to bright careers. But we really don't know whether they succeeded or failed, or whether the program had anything to do with this success or failure. Without continuous or appropriate evaluation we go from one whim to another without knowing whether the products of our earlier efforts were highly successful or complete failures.

Even when some components of an evaluation system are operative, others may break down. We recently observed a splendid program effort in evaluating the student teaching program of a major university. This university placed 70% of its student teachers in school systems remote from campus. Resident university supervisors were employed to assess student teacher peformance to determine the strengths and weaknesses of specific programs through measures of student teaching competency. They employed interaction analyses, pre-, mid-, and post-experience inventories, surveys of students and cooperating teachers and administrators, extensive observational data, diagnostic instruments, cognitive level assessments, and a host of other devices. But then we found that none of this information was ever referred back to the teacher education departments! Just imagine that! Once students had completed their work in subject matter or pre-clinical specialties, their respective departments never heard from them again. This was the only evaluation effort the institution employed, yet when faculty were questioned about the program they bragged of its quality, citing as evidence the small number of informal complaints received about their products. Even the supervisors drew unfounded conclusions regarding the academic program. They assumed that if student teachers were successful, the academic program must be functioning at an optimal level. Because this institution served a wealthy and well-educated clientele, however, it was just as plausible that graduates were successful in spite of their experience in the program. Or perhaps their skills could have been improved. We will never know.

I owe to James Raths of the University of Illinois a debt of gratitude for pointing out another mountain against which evaluation efforts are frequently dashed. Professor Raths notes that teacher educators frequently get trapped into assuming that the assessment of individual teaching competencies is the essence of program evaluation. Says Raths:

> It is a matter of getting involved in an infinite regress. If you respond by saying we want good programs to have good teachers, I could ask why we want good teachers. The answer is that we want to have good products graduate from our schools. If we pursue this questioning, we could ask why we want outstanding school graduates; the answer is to have a better society. Why do we want a better society? So that...where does it end?

Raths' best guess is that "...each of us has an arbitrary focus on an infinite chain." The focus of program evaluation, he believes, "is properly on the quality of programs and not on the quality of teachers."* Teacher educators, however, are often pressed to use program evaluation in lieu of the assessment of individual skills. This confusion of apples and bananas has been greatly abetted, and even enshrined into law, by state program approval efforts. Many state departments of education have found a cheap way to license teachers by handing a certificate to anyone having completed an "approved" or accredited program. Yet no model of program approval or accreditation can possibly determine the competencies of individual products of those programs. It is not a task appropriate to program evaluation, nor for that matter to national accreditation, although we frequently are criticized for not doing it. Even if we could, the impact of school systems upon the behaviors of individual teachers probably would overshadow program effects within a few short years. Programs are designed as vehicles for providing opportunities for behavioral change. There can be no guarantee that these changes will last or will not be altered through subsequent experience.

---

*James Raths. Personal correspondence of February 25, 1980.

The Evaluation of Graduates

Nothing herein is meant to detract from the importance of evaluating the products of our programs. This is not to say, however, that program evaluation should rest exclusively on the evaluation of graduates. Nor should we assume that evaluation of graduates and assessment of individual competencies have the same purposes or aims.

Because so much program evaluation has focused on the evaluation of graduates, there has accumulated a significant amount of information on the subject. We find, of course, that in the area of graduate assessment have occurred our greatest number of mishaps and total disasters. Use of program evaluation in an attempt to guarantee individual competency has resulted in many institutions attempting to evaluate the entire population of their graduates, with an almost universal lack of success.

NCATE published in 1970 the first revision of its standards that required institutions to evaluate their graduates and use the results of these evaluations for program modification and long-range planning. The Council's intent was to complete the loop of a system by providing feedback on program output in conjunction with other evaluation data that could be used to modify input and process variables. We now have a ten-year history of efforts to evaluate programs, providing a reasonable basis for drawing some conclusions about the efficacy of teacher education evaluation on a national scale. We regret to report that these efforts have been less than impressive.

For the past two years violations of standards of program evaluation have constituted the top two reasons for denial of NCATE accreditation for both basic and advanced programs. Even where denials did not occur, lack of effective evaluation systems led the list of weaknesses cited for baccalaureate and graduate programs at 143 institutions reviewed.

Where programs failed there generally was no lack of motivation. Evaluation of graduates and other aspects of progam assessment have become the most common topics at seminars, meetings, orientation sessions, and anywhere else accreditation concerns are raised. We recognize that in some instances technology of measurement simply has not risen to a level consistent with professional expectation. Yet with that concession notwithstanding, we have observed many common pitfalls that almost guarantee the failure of efforts to evaluate graduates.

Institutions that concentrate their evaluation activity on the assessment of graduates frequently tack this evaluation on to preparation programs almost as an afterthought. Rather than construct the concept of evaluation of graduates into the general evaluation system of a program, many institutions pay little attention to evaluation until they begin to prepare for an NCATE evaluation or other program review. Some institutions regard program evaluation and national accreditation as one and the same thing. (One result of this has been that institutions attribute great cost to NCATE accreditation because they add every nickel and dime spent on evaluation to the total accreditation bill. The difficulty in countering this claim is that the only reason many schools do conduct evaluations is because of accreditation. Few successful program evaluations are conducted when the only motivation is to "pass professional muster.")

The least productive approach to program evaluation has been that of evaluating graduates through attempted surveys of entire graduating classes. Institutions have demonstrated a fundamental lack of ability to track graduates and sustain contact with them. This has resulted in survey returns of 40% being considered phenomenal, with normal returns frequently running at less than 15%. Low return rates breed problems of sampling. One must ask who are the respondents? Those who are unemployed, angry at their former institution? Those who have enough time to write nasty responses? Or are they the extremely dedicated who enjoyed their program and would not criticize it even constructively?

What questions can one ask that will provide useful information? Surveys must measure what they can, not what may actually be there. The result is that graduates are asked to respond to mundane and nebulous questions: "What did you feel was the most important aspect of your training program?" "What did you feel you missed in your preparation program?" Answers to such questions are predictable and less than helpful. Individuals respond that the most important aspect of their program was student teaching. They usually indicate that they needed work in discipline. These responses tend to occur even when individuals have had inordinate amounts of student teaching and when classroom management and discipline were primary components of program experience. We find that surveys often measure not the knowledge and skills students acquired in the program, but what common wisdom tells them to think they acquired, or did not acquire. Surveys tell us more about peer knowledge and group methodology than they do about preparation programs completed by prospective educators.

Surveys of employers have also proved less than helpful. One midwestern institution recently became suspicious when for the third consecutive year certain employees of their graduates responded with almost identical comments of praise for program graduates. Neighboring institutions were therefore contacted to determine the nature of responses they were receiving from these same employers. Some were amused to find that the same principals and superintendents who had indicated they would hire no graduates other than those of a particular institution, had told competing institutions the very same thing. Although exceptions do occur, in general employers seemed unwilling to be candid about the quality of personnel they hire at least when responding to mailouts. Perhaps to be critical would be to criticize their own selection and employment procedures. Some may actually fear legal reprisals from criticized employees.

Another common problem with surveys is that they attract information that is not current. Many schools report that when they do receive constructive information from respondents, it is outdated. Former students criticize professors who have not been employed for years and courses that no longer are in the curriculum. They typically advise that coursework in reading and special education should be added when it probably became a part of the curriculum years ago. Survey techniques tend to ask simple questions and evoke simple responses--gut level reactions that may be no more timely than they are accurate. Surveys do not get at classroom technique, level of knowledge of the teaching speciality, or anything else that one would normally list as high on the priority of essential knowledge and skills for educators.

Evaluation efforts often fail because they are insufficiently financed. The adage that one gets what one pays for could be no more true than in reference to the assessment of graduates. A last-minute, one-page survey questionnaire will glean about as much useful information as the time and money that go into such minimal efforts. The last place one wants to become stingy is in the area of evaluation. If done appropriately, evaluation can actually save money by making programs more cost effective, to say nothing of improving general academic quality.

It is not my assignment to describe successful program evaluation activities and techniques. But with respect to the evaluation of graduates, I will note that NCATE has found the most successful evaluation systems to be those that employ sampling, sustain contact with a limited number of carefully-selected, representative graduates, and that place evaluators in the field, rather than rely exclusively upon surveys. All too frequently, institutions completely overlook such face-to-face techniques as interviews of graduates, their employers and students. Sampling the opinions of a diverse population of graduates, their peers and employers, can provide terrific insights into program needs, not only for pre-service preparation, but for in-service programming as well. Many institutions overlook the potential of internships and other residency experiences that could provide opportunities for acquiring invaluable feedback on the performance and capabilities of graduates and those who supervise them. Argyris and Schon (1977) have demonstrated that some residency or internship for teachers and school support personnel would go far in providing invaluable data on program strengths and weaknesses.*

---

*Chris Argyris and Donald Schon, Theory in Practice: Increasing Professional Preparation. New York: Jossey-Bass, 1977.

Using Results of Program Evaluation

As common as any problem with evaluation efforts has been a general inclination of professional faculties to ignore results of program evaluations. Even in some nationally-acclaimed efforts, scant use of information resulting from these evaluations has been made by resident faculty. One cannot expect merely because an evaluation effort bears fruit, that teacher educators will beat down the door to acquire criticisms that might suggest need for program modification and change, reordering of departmental priorities, or--God help us--reassignment of faculty.

If an evaluation program is, in the final analysis, to be ultimately successful in modifying faculty behavior, then from the outset it should involve program faculty in articulating evaluation efforts with program aims and objectives. To do less is to guarantee failure of the evaluation mission.

Again, program evaluation should be comprehensive, examining the curriculum, procedures, and processes that affect student admission, courseware, clinical and field experiences, as well as the evaluation of graduates. Institutions that fail to assess the function or quality of any of these components clearly limit not only the success of their evaluation effort, but also severely restrict the probability of academic success.

Summary

To summarize, here are 12 suggestions of what not to do if one aspires to establish a successful program evaluation effort:

A DOZEN DON'TS

1. DON'T fail to evaluate program entrance variables, including admission criteria, counseling procedures, and other factors that may affect operations throughout the entire program.

2. DON'T fail to evaluate the criteria and procedures of program retention, including assessing the effectiveness of screening criteria, procedures for counseling out the unsuited, and for determining remedial procedures for those whose performance is minimal.

3. DON'T expect evaluation of professional programs to result in tight statistical data that can be interpreted with little judgment. Professional decisions are subjective ones.

4. DON'T place all your eggs in one evaluation basket. Diversify. Provide checks and balances for the testing of validity throughout the program.

5. DON'T expect program evaluation efforts to be successful if conducted only sporadically. Evaluation should be an integral part of each aspect of the program from planning to execution, from admission to graduation.

6. DON'T try to get by on the cheap. You get what you pay for in program evaluation. Evaluation efforts usually are cost effective. Their primary value is that they spotlight areas in which fiscal and personal resources are being wasted, as well as those components that are functioning efficiently.

7. DON'T fall into the trap of assuming that the assessment of the individual skills of graduates is equivalent to the evaluation of programs.

8. DON'T bother with correspondence surveys of employers or graduates, unless this is the final resort.

9. DON'T fail to use interview techniques. Concentrate opportunities afforded by practica and internships for acquiring face-to-face reactions to questions regarding program operation.

10. DON'T ignore important evaluation feedback even if it is critical and politically unpopular in the short-run. Data supporting negative connotations can best be handled when kept in the context of long-range planning.

11. DON'T assume that just because evaluation data are available that individuals will automatically use it to modify and improve programs. A conscious effort must be mounted to school faculty and administrators into employing evaluation data in planning and program modification.

12. DON'T treat evaluation as an afterthought or conduct program assessment because some state department or accrediting agency requires it. An evaluation should be as much a built-in component of professional programs as curricula, clinical experiences, or any other fundamental part.

USE-ORIENTED EVALUATION

James R. Craig
Ronald D. Adams
Western Kentucky University

INTRODUCTION

Evaluation of teacher education programs has been a topic of intense discussion both inside and outside the educational community in the last decade. The focus to a large degree has centered on practical considerations related to data collection and analysis methodologies appropriate for use in assessing the effectiveness of special public school programs, federally supported projects, and the like (e.g., the classic Campbell and Stanley (1963) text). Also during this period of time, a great deal of journal space has been devoted to the theoretical aspects of evaluation in teacher education, such as formative and summative evaluation (Scriven, 1967) and goal-free evaluation (Scriven, 1972), as well as models of the evaluation process (Popham, 1972; Wortman, 1975; Stufflebeam, 1978). In all instances, the emphasis has primarily been on method.

In the past decade, however, many individuals involved in evaluation have become increasingly aware that evaluation is not "method bound" but rather is bound by the social/political context within which evaluations are conducted and data are used (Weiss, 1972; Suchman, 1972; Mathis, 1980). Ideally, evaluation data are used to make objective program decisions regarding program development, modification, operation and the like. However, it often seems rare that evaluation data are used in such a direct, straightforward manner. Therefore, one of the primary evaluation concerns in recent years has been the development of ways to obtain more effective use of evaluation data in program decision making (e.g., Patton, 1978; Mathis, 1980). The important term is use.

The definition of use typically advanced is that use occurs when evaluation data are directly employed in objective, observable ways in program modification and operation (Cohen, 1977; Patton, 1978; Mathis, 1980). However, such a definition does not fully incorporate the social/political context in which programs operate and the fact that evaluation data constitute just one of several bits of information that entered into program decision making in teacher education (e.g., governmental statutes and personal gain). A more realistic view is that use of evaluation data in program decision making is almost always diffuse in nature and not always (if ever) directly observable (Patton, 1978); use is an iterative process which focuses on the assimilation of evaluation information into the decision making process.

How then can evaluation systems be structured to better use the data which the systems provide? We believe the answer to that question is to develop evaluation systems for use in teacher education where the emphasis from the very beginning is on building mechanisms to increase the systematic inclusion of evaluation data in program decision making. The evaluation system we intend to present for your consideration focuses on use and was developed out of both our own attempts to conduct evaluations of programs and in trying to assist others. Many of the ideas are not new; most, if not all, have been directly suggested or at least anticipated in one way or another by others who are actively involved in evaluation. What is new is the emphasis on use and the manner in which that emphasis is incorporated into teacher education evaluation systems.

BASIC PREMISES

Our proposals are based on the following four premises:

1. The purpose for establishing an evaluation system for a teacher education program is to provide relevant, meaningful information to be systematically incorporated into program decision making. Establishing an evaluation system simply to have one in place for an NCATE visit cannot and should not provide the sole motivation. The NCATE standards are well intended, but it remains for us, the professional teacher education community, to give these standards meaning and value--especially regarding evaluation of our own programs.

2. All individuals responsible for program implementation and operation should be identified. The organizational structure within which most teacher

education programs operate and are evaluated can be represented by the hierarchy shown in Figure 1. The chief administrative officer (usually a dean or department head) is the individual responsible for establishing the parameters (e.g., the allocation of resources) within which programs must operate. Also contained within the typical organizational structure is a program administrator (e.g., a department head or a coordinator) charged with operating the program, a program staff (e.g., a departmental faculty) responsible for delivering program services, and program participants (e.g., undergraduate students) who are the individuals who receive the services the program was created to render.

3. If evaluation data are to be effectively used in program decision making, then individuals at all levels of a program must be involved in the evaluation effort. This means that not only program administrators and program staff must be included in the development and operation of the evaluation system, but that administrators higher in the organizational structure must also actively participate in the evaluation process, incorporating the evaluation data into their own decision making regarding programs and related policy concerns.

4. The evaluation of teacher education programs must operate within the organizational framework which currently exists. Furthermore, evaluation should be a part of the ongoing operation of the program; it should not be perceived as a "lay-on" by administrators, legislators, or other "outsiders." In addition, it should be realized at the outset that it is very unlikely that new monies will be made available to fund program evaluation efforts. Evaluation will have to occur within the context of the operation of on-going teacher education programs, planned professional development efforts, and related activities.

USE-ORIENTED EVALUATION

The use-oriented evaluation system we propose is composed of five components (see Figure 2): Initiation and Verification, Planning, Implementation, Data Collection, and Feedback and Utilization. These components are each discussed below with the focus being what can be done to increase the frequency of use of evaluation information in program decision making.

Initiation and Verification

A summary of the organizational chart as it reflects the activities at both the Initiation and Verification and Planning stages is presented in Figure 3.

The impetus for implementing an evaluation system may come from outside or from within the organization. Regardless of the source or the position of the initiator, the Initiation and Verification phase of an evaluation ultimately must begin with the chief administrative officer responsible for the program unit--usually a dean. Without that individual's understanding of, commitment to, and approval for the evaluation of a program, the evaluation effort will not be successful.

Individually (or perhaps with the assistance of an experienced evaluator), the chief administrative officer should conduct a preliminary overview of the evaluation and an analysis of its various ramifications. The intent is to identify the chief administrative officer's perceptions of the present institutional and program circumstances, the ideal program, the need to conduct an evaluation, the possible options for implementing the evaluation, the possible program implications that might be suggested by different evaluation data, and the resource restrictions within which the evaluation must be conducted. The objective is NOT to have the chief administrative officer establish the form and substance of what the evaluation should be, but rather to provide her/him with an understanding of how the evaluation process may operate. This process will allow the administrator to realistically set the parameters within which the evaluation must function (e.g., the budget), develop a commitment to the evaluation process, and identify possible program decisions which could result.

If, at that point, the decision is still to institute an evaluation system, then the preliminary review should be repeated with those individuals within the organization who serve as the primary decision makers for the programs to be evaluated. The decision making team should be identified by the chief administrative officer. It should be composed of individuals who collectively facilitate the establishment of specific program goals, functions, and operational procedures, as well as those who are responsible for decisions regarding program modification and continuance (e.g., an assistant dean, a department head, and a program coordinator). The same considerations addressed by the chief administrative officer should be addressed by the decision making team. In addition, this group should determine the procedures for the creation of the planning and evaluation team.

Planning

The planning and evaluation team is the primary initiation and operational force in an evaluation effort. The team determines the details of the evaluation (e.g., the types of data to be collected) and is responsible for communicating to all concerned the form and substance of the various evaluation activities. So that the team can function effectively, it should be composed of both formal and informal program leaders and be limited to a maximum of ten members. The exact composition of the team will be specific to each situation and determined by such things as the number of program staff, the leadership style of the chief administrative officer, budget restrictions, and similar factors.

The planning and evaluation team has four charges. First, it should create a viable, flexible, workable evaluation plan agreeable to the team members that includes at least the following characteristics:

a. a specification of the evaluation data required to make decisions regarding the present status of the program (within the boundaries established by the decision makers);

b. the rated importance of the various data consistent with the teacher education program goals and objectives and the current knowledge regarding effective teaching;

c. an identification and listing of the possible sources of data;

d. an evaluation of the possible data sources and collection procedures in light of access, cost, developmental time, staff development needs, and time delays;

e. a prioritization of the possible data sources and collection procedures in terms of the evaluation data required and related resource restrictions;

f. and, a description of the available data collection procedures selected and/or adapted, new procedures developed, and the data collection training required (if any).

The second charge of the planning and evaluation team should be to create an effective, operational communication system. The communication system should be used during the development of the evaluation plan to ensure the involvement of all individuals in planning and modifying the evaluation system. If implemented properly, the system should function after the evaluation plan is operational to disseminate evaluation data to all concerned faculty and administrators for use in program decision making.

The third charge of the team should be to designate someone to be the evaluation manager. Space does not permit us to fully elaborate the role of the evaluation manager. At this point it is sufficient to say that the evaluation manager should serve as the team's executive officer and be responsible for the "nuts and bolts" implementation and operation of the evaluation plan. The person probably should have the needed technical skills (although these could be acquired through training) and should be a capable group facilitator.

The fourth charge is that the planning and evaluation team and the evaluation manager should identify and establish an advisory group consisting of field based educators. Input from the practicing arm of the teaching profession is necessary to gain the perspective of the school administrators who will ultimately employ graduates and the perspective of the practicing teachers who can provide expertise gained from on-the-job experience. Others who may be important members of an advisory team would be university educators in areas other than teacher education (e.g., content areas and data processing).

Implementation and Data Collection

Implementation and Data Collection reflect the activities associated with the operation of the evaluation plan. Briefly, the evaluation plan established by the planning and evaluation team should be specified to the extent that procedures for data collection are established to include selection of instrumentation and data collectors, procedures for selecting participants, and identification of data management systems. The evaluation plan is then implemented and operated by the evaluation manager under the direction of the planning and evaluation team.

Feedback and Utilization

One of the most important aspects of the evaluation system is the development of an active, effective communication system that is continually operative from the very beginnings of the evaluation effort. It is extremely important to realize that the generation, dissemination, and utilization of evaluation data is NOT a lock-step, linear sequence (Havelock & Lindquist, 1980). As was noted earlier, evaluation of teacher education programs is conducted within a social/political environment. Therefore, the key to the successful operation of an evaluation program is continual involvement of all individuals concerned through an ongoing exchange of ideas and values coupled with the evolution of a commitment to make the educational program the best it can possibly be. The form that any particular communication system may assume in a given instance will be peculiar to that situation. Regardless of the particular system that evolves, the emphasis should be on the effective use of evaluation data in program decision making. As evaluation data become available, the data should be systematically reviewed, assessed, and interpreted by all those involved. In this regard, there are several important considerations. First, the evaluation data should be analyzed and presented in a form that is both consistent with the program objectives and readily understood by the users of the information. Second, there must be a formalized process by which these data are reviewed by all staff and applied in making program decisions. Third, the change process should be documented and the evaluation system reviewed in light of any changes made.

CONCLUSIONS

We believe that involvement and "evolvement" are the keys to the development of evaluation systems where evaluation data are effectively used.

First, policy makers, administrators, and program personnel at all levels must be actively involved. They all should participate in the determination and operation of an evaluation system if it is to be one where the data are validly collected and analyzed and one where the data are seriously reviewed and considered in making programmatic decisions. Second, it should be realized that evaluation must be viewed as iterative in nature where both the evaluation processes and products evolve over time. The system should reflect current concerns but also be flexible enough to adapt to new and changing needs. If we want evaluation systems for our programs to be something more than decorative ornaments, that is, if we want them to provide useful information for program decision making, then we must position them in the forefront so that we might adapt them for their most effective and efficient use in meeting the needs of the education profession we collectively serve.

References

Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Cohen, L. H. Factors affecting the utilization of mental health evaluation research findings. Professional Psychology, 1977, 8(4), 526-534.

Havelock, R. G. & Lindquist, J. A conceptual framework for increasing the impact. In J. Lindquist (Ed.), Increasing the impact of social innovations funded by grantmaking organizations. Battle Creek, MI: The W. K. Kellogg Foundation, 1980.

Mathis, W. Evaluating: The policy implications. Phi Delta Kappa CEDR Quarterly, 1980, 13(2), 3-6, 22.

Patton, M. Q. Utilization-focused evaluation. Beverly Hills, CA: Sage, 1978

Popham, W. J. An evaluation guidebook. Los Angeles: The Instructional Objectives Exchange, 1972.

Scriven, M. The methodology of evaluation. In R. W. Tyler, R. M. Gagne, and M. Scriven (Eds.), Perspectives of curriculum evaluation. Chicago: Rand McNally, 1967.

Stufflebeam, D. Meta evaluation: An overview. Evaluation and Health Professions, 1978, 1, 17-43.

Suchman, E. A. Action for what? A critique of evaluative research. In C. H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn and Bacon, 1972.

Weiss, C. H. Evaluating research: Methods for assessing program effectiveness. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Wortman, P. M. Evaluation research: A psychological perspective. American Psychologist, 1975, 30, 562-575.

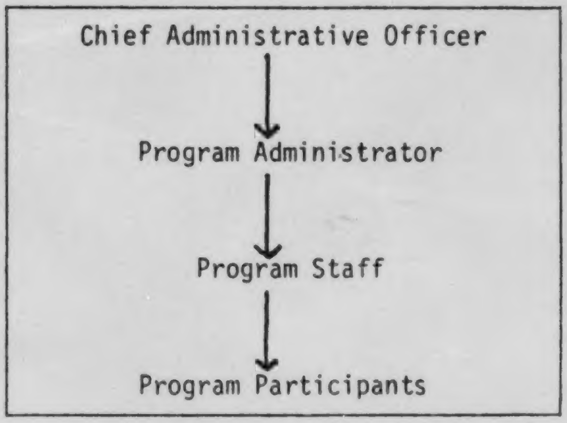

Chief Administrative Officer
Program Administrator
Program Staff
Program Participants

Figure 1: A Typical Organizational Structure

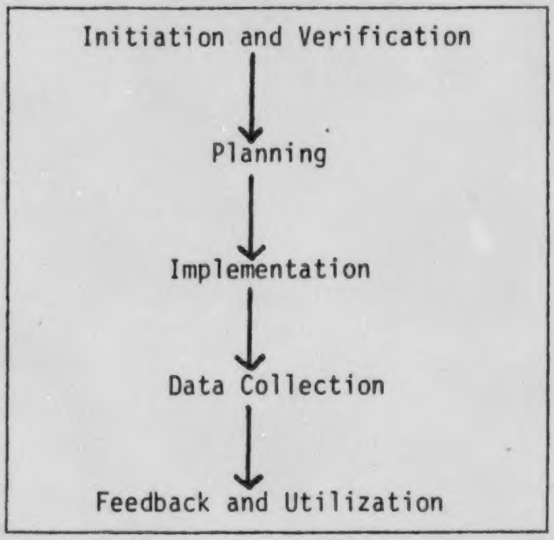

Initiation and Verification
Planning
Implementation
Data Collection
Feedback and Utilization

Figure 2: An Evaluation System

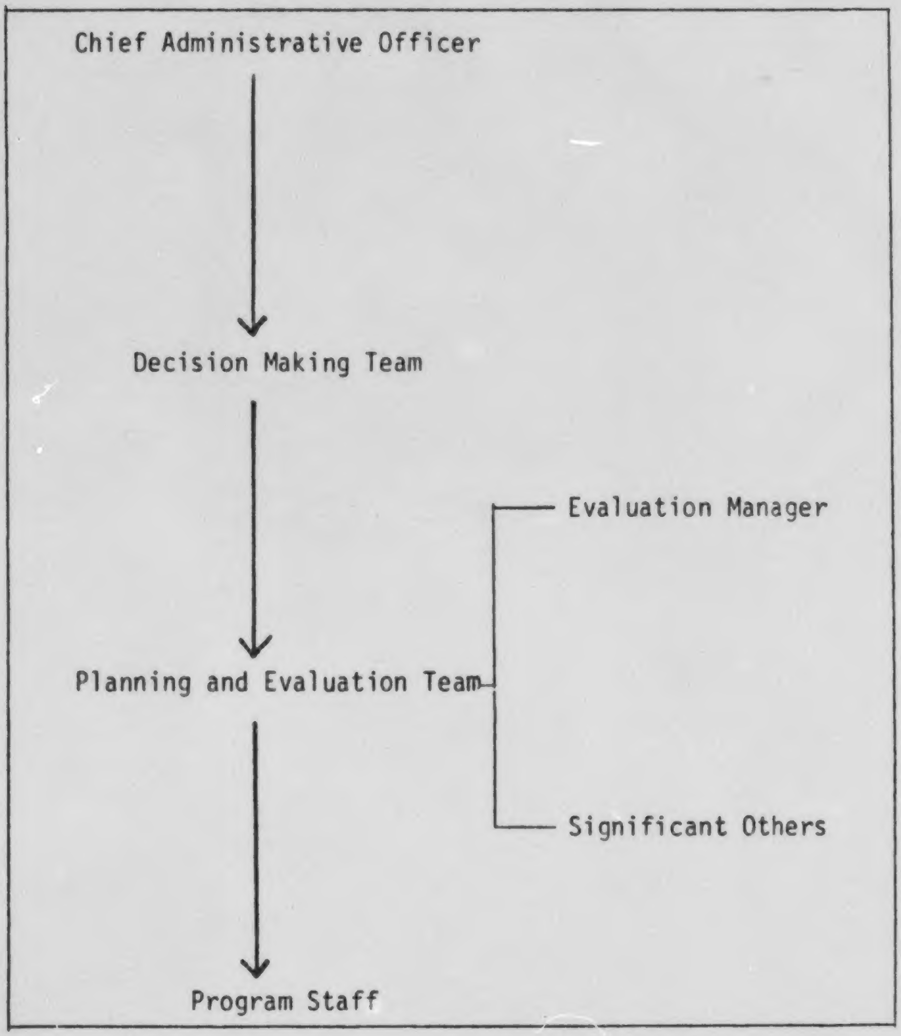

Chief Administrative Officer
Decision Making Team
Planning and Evaluation Team
Evaluation Manager
Significant Others
Program Staff

Figure 3: Use Oriented Evaluation Sequence

DESIGN CHARACTERISTICS FOR MEANINGFUL TEACHER EDUCATION EVALUATION

Jerry B. Ayers
Tennessee Technological University

Evaluation of teacher education program graduates has been one of the most difficult problems facing institutions of higher education during the past decade. Various institutions have attempted to develop models for evaluating the product, i.e., graduates of a teacher education program; however, only a limited number of institutions have implemented a comprehensive model. Western Kentucky University (WKU) and Tennessee Technological University (TTU) are two institutions that have developed and implemented models for the evaluation of teacher education graduates. The models serve as a guide for gathering evaluative data that can be combined with other information for the purpose of evaluating the teacher education program. The institutions have been employing these models for over eight years for the purposes of curriculum development and improvement. The models are still incomplete and inadequate; however, much information has been gained about evaluation of teacher education graduates, and a definite set of characteristics have evolved which can form the basis for the development and implementation of teacher education follow-up evaluation models by other institutions.

Design Characteristics

Ten major design characteristics are essential for a meaningful program of teacher follow-up evaluation. This section presents an outline of each of these characteristics, some of the problems that might be encountered if the characteristics are not considered, and a summary of how the charactertistics have been made a part of the models employed in follow-up studies of the graduates of the teacher education programs of WKU and TTU.

1. The faculty, administration and governing body of the institution must be committed to conducting a program of teacher evaluation.

If an evaluation system is developed in a vacuum without adequate input from the faculty, the results of the evaluation effort may be lost. The faculty are the primary consumers of the results of a program for the evaluation of teacher education graduates and if they are to make use of the data and results from the study, they must have adequate input in order to answer questions they have relative to program development and improvement. In a similar manner, the administration and governing body of an institution must be involved in the project, in order for them to understand and provide the needed resources for conducting the evaluation program. Without adequate involvement and commitment on the part of the faculty, administration and governing body of an institution, the evaluation program cannot succeed.

At WKU and TTU, the initial impetus for the development and implementation of the models came from efforts to meet the evaluations standards of the National Council for Accreditation of Teacher Education (NCATE). However, the faculty, administration, and to some extent the governing bodies of the institutions were involved in the program of evaluation from the very beginning. In this manner, the faculty and administration felt the project was a part of their efforts in program development and improvement.

2. A teacher evaluation program must be viewed as a part of the total teacher education program and not as an isolated project.

Industry spends a significant portion of its resources on product evaluation. All industries in the competitive market-place have systematic, valid procedures for testing how well its products perform the functions for which they were designed. Institutions of higher education are only beginning to realize the value of the industrial model for the evaluation of its product, i.e., the graduates of its programs. It is realized, however, that the application of the industrial model is difficult when dealing with people. Teacher follow-up evaluation must be integrated into the total program of teacher education. Sufficient resources must be allocated to the evaluation efforts and they must be made a part of the total teacher education program. If a teacher evaluation program is carried out in an isolated manner, the results will not find their way into the total program for the preparation of teachers. The evaluation efforts would thus be lost for purposes of program development and improvement.

WKU and TTU have had systematic programs of teacher follow-up evaluation for over eight years. At each institution individuals have been assigned the responsibility of conducting the evaluation efforts and sufficient resources have been allocated for carrying out the evaluation studies. The administration of each of the institutions supports the evaluation efforts and have made every effort to provide the needed resources. The teacher evaluation efforts have been built into the total program for the preparation of teachers.

3. A teacher evaluation program must be specific, yet reflect the total teacher education effort. The evaluation efforts must reflect the goals and objectives of specific programs and the overall program for the training of teachers.

Training programs include general knowledge which is common for all prospective teachers, and specialized knowledge which is applied according to the level or area of teaching. Most institutions train prospective elementary teachers, secondary teachers in a variety of fields, and possibly special education teachers. Evaluation efforts must include not only elements common to all teacher training but to training for work in a specialized field. For example, the training that prospective teachers receive in methods of teaching varies from elementary to secondary programs of study. This variation must be reflected in the evaluation program that is established. Also, within a teacher training program there are certain common elements which must be evaluated, for example, basic education requirements which include competency in such areas as basic psychology, social foundations of education, and history and philosophy of education.

The specification of objectives for the training of teachers is an inherent element in an adequate evaluation system. If the plan for the evaluation of teachers is not based on both the specific and general objectives of the teacher education program, the evaluation may turn into a useless exercise. The results of the evaluation effort would not be tied to any specific effort that may have been made in the teacher training program. The instruments used by WKU and TTU to collect data reflect the specific and general objectives of the programs for the preparation of teachers at the institution.

4. An evaluation program must be continuous and longitudinal in nature.

In order for a program of teacher evaluation to be meaningful, it must be continuous and should be longitudinal. For effective program improvement, there needs to be a continual input from the graduates of the program. As was pointed out earlier, industry uses a continuous model for evaluation in order to change and improve its products. The analogous situation holds for programs which train teachers. Programs of evaluation which are not continuous neither allow for measurement of changes that may take place in a program, nor for determination of developing problem areas.

The models that are in use at WKU and TTU have been in continuous operation for over eight years. In the case of the model used at WKU, data have been collected from graduates during student teaching and their first, third and fifth year in the profession. Data have been collected each year from the first through the fifth year for graduates of programs at TTU. A new group of graduates enter the evaluation system each year. It is felt that the systems which are in place will continue on an indefinite basis.

5. An evaluation program must provide for and reflect a knowledge of the state-of-the-art in educational research and evaluation as it relates to teacher education.

The state-of-the-art of teacher evaluation is in a continual state of change. Since the early sixties, many resources have been allocated to the study of effective teaching, and in the seventies a number of individuals and institutions developed models and plans for the effective evaluation of teachers. If an institution is going to develop and operate an effective plan for the evaluation of teachers, it must constantly be aware of the research-based changes which are being made in evaluation. It should be pointed out that NCATE Standard 6.1 states, "The institution keeps abreast of emerging evaluation techniques and engages in systematic efforts to evaluate the quality of its graduates upon completion of their program of study and after they enter the teaching profession" (Standards, 1979, p. 11).

WKU and TTU have regularly monitored the research literature related to teacher evaluation. The basic models that were developed nearly a decade ago are still being used; however, changes have been made in the instrumentation and other aspects of the programs that reflect the results of research on teacher evaluation. These two institutions have maintained continual contact for the purpose of sharing information. Recently, through the efforts of the Research and Development Center for Teacher Education at the University of Texas and others, a network of individuals interested in teacher evaluation is being formed (Hord & Hall, 1978).

6. An evaluation program must reflect thorough planning prior to implementation.

The implementation of an evaluation program within an institution can be a difficult task. Adequate planning prior to implementation is essential if the objectives of evaluation, i.e., program development and improvement, are to be recognized. Following are six important points that must be considered in planning and implementing a program of teacher evaluation.

a. Cooperation of schools

Prior to the initiation of any type of study of the graduates of a teacher education program who are practicing teachers in the schools, it is essential to obtain the permission and consultation of school administrators. Also, it may be necessary to consult with and have the permission of local teacher negotiating groups. For example, in some areas of the country extensive follow-up evaluation in the schools has not been possible because the local bargaining group for the teachers did not make provision for such activities in the schools, or would not permit such activities since they might infringe upon the system of evaluation already in use in the schools. Adequate permission for data collection in the schools is essential if a project of teacher evaluation is to be a success. Without such permission and a full understanding on the part of the school administrators of the use of the data which is to be collected, the project cannot succeed.

b. Subject selection

Various methods exist for the selection of subjects, taking into consideration the constraints that were imposed in (a), above. The basic premise of subject selection is that the process must be a defensible system and should be the same from year to year. It appears that the most practical system of subject selection, even though it is not the most defensible from a scientific point of view, is to select teachers from a pool of those graduates who volunteer to participate in the project.

c. Data sources

Evaluation of teachers can be approached from several standpoints. Data collection should not, however, be limited to only one source (for example, only the graduate). Frequently, in the past, institutions have sought evaluative data by simply asking the graduate to rate the value of certain experiences that were a part of their teacher training program. Although this is an inadequate source of data, it is an important source. Data for an adequate evaluation program should be collected from a variety of sources, including the teachers, the teachers' supervisor and peers, their students, and probably most importantly, by independent observers in the classroom.

d. Qualitative and quantitative

As an integral part of any evaluation effort, consideration must be given to both qualitative and quantitative data collection. Quantitative data can readily lead to analyses allowing for the use of inferential statistics. However, in many cases qualitative data can be more meaningful and useful to faculty and administrators in program development and improvement.

e. Points of data collection

The points of collection of data in a teacher evaluation program must be planned prior to implementation. Various plans have been used in the longitudinal studies that have been reported in the literature. The most common systems which are in place include data collection during student teaching and the first, third and fifth year after entering the teaching profession, or data collection during the first year after entering the teaching profession and each year thereafter for up to five years. Little useful data with regard to program improvement and development can be collected after five years because of a number of factors. It is essential that data be collected during the first year after entering the profession and for at least one or more years up to five years. If such a plan for data collection is not initiated at the start of the evaluation project, much useful data may be lost.

f. Data collection, instrumentation and personnel training

Initially a system for data collection must be developed. As a part of this system, instrumentation must be selected or developed which can be used to collect quantitative and qualitative data that will measure the achievement of the objectives of the preparation program of the teachers. Numerous standardized instruments are in existence; however, it may be to the advantage of the institution to develop instrumentation that will better meet its particular needs. As an integral part of the development of the plan for instrumentation and data collection, personnel who will be used to collect data must receive special training. For example, if a system of interaction analysis is to be employed to collect data in the classrooms of the teachers, then the research assistant or data collector must receive adequate training in order to collect valid and reliable data.

Throughout the life of the evaluation project, instrumentation and methods of data collection must be systematic. For example, if one type instrument is used one year and the following year a second type of instrument is used, data from the two data collection periods may not be compatible. Thus, one or more years of useful data may be lost in a longitudinal study. Instrumentation and a plan for data collection must be developed early and must stay in place for several years. Also, personnel must receive adequate training each year, in order for data to be compatible from year to year.

g. Comments about planning

If any one of the above constraints relative to implementation of a teacher evaluation program are not followed, the results of the project may be suspect and will be of little use for program development and improvement. WKU and TTU each have developed detailed plans for the implementation of their follow-up studies. The plans are rather detailed and space does not permit a lengthy description of the projects at this point. The reader is referred to Teacher education program evaluation and follow-up studies: A collection of current efforts (Hord & Hall, 1978) for a more complete description of the implementation of the projects or the College of Education of the two institutions.

7. An evaluation program must reflect the most efficient use of resources for data processing, storage, retrieval and analyses.

Prior to the start of a teacher evaluation project, an institution must inventory its resources relative to those available for data processing, storage, retrieval and analyses. A decision must be made at the start relative to the amount of data to be collected and stored and how it can be retrieved. Access to a high speed digital computer is essential if a large scale project is attempted. For example, one institution attempted to replicate the teacher evaluation model that is in use at WKU. The institution found that after collecting data for one year on a sample of 30 teachers, the computer available on their campus did not have adequate storage space available. As a result the data analyses had to be carried out in a cumbersome manner resulting in the need for additional resources.

WKU and TTU have been fortunate in having access to large scale computers with adequate storage. The problems of data storage, analyses and retrieval have been minimized. It should be pointed out that both institutions have had serious problems with data management and control. These problems have been due to the large amount of data collected each year and kept in continuous storage for use in making comparisons across years and for special analyses.

8. An evaluation program must provide for an effective communication system for input into the evaluation process and feedback of evaluation information.

There must be an effective system of communication for input into the project from the faculty who will eventually use the data and results of the evaluation program. In turn there must be an effective system for communication to the faculty such that the results can be used in program development and improvement. The most efficient system for providing faculty input into the project is to make certain at the outset that the faculty feels it is their evaluation program and therefore their input is of value and use. It should be pointed out that NCATE Standard 6.2 states, "The institution regularly evaluates its teacher education programs and uses the results of its evaluations in the modification and improvement of those programs" (Standards, 1979, p. 11). Without an adequate system of feedback into the program for the preparation of teachers, the latter part of the Standard cannot be met.

WKU and TTU have developed programs for the feedback of information into the programs for the preparation of teachers. Both institutions use a systems approach for the operation and evaluation of their programs. Thus, there are natural built-in feedback loops. On the more practical side, the institutions prepare reports on a regular basis that are made available to faculty and administrators, as well as special reports and papers relative to specific areas and studies.

9. An evaluation program must provide for a workable system of evaluation data utilization in program decision making.

The success of an evaluation program may be measured by the use that is made of the results. If the results from the program are not used in a meaningful manner, then the evaluation program may be a total failure. Dissemination of the results of the evaluation program may not be sufficient to make changes and needed improvements in a program for the preparation of teachers. The institution must have a built-in system for the utilization of the results of the evaluation program. As was noted under (8), a systems approach to the training of teachers can have a built-in subsystem or feedback loop that will make possible the use of the results of a teacher evaluation program.

10. An evaluation program must have a component to assess the effectiveness of the evaluation process.

Industry constantly evaluates the standards and evaluation procedures that it employs in the evaluation of its products. Any evaluation program should have a built-in system for assessing its effectiveness as a tool for gathering data and making meaningful decisions. Therefore, the program for self-assessment may take the form of the external review by consultants or an internal review by the project staff. Such a subsystem within the evaluation system should be developed at the beginning of the project. WKU and TTU have developed systems for review of their evaluation models.

Summary and Recommendations

This paper has described ten characteristics that must be taken into account in the development of an adequate system for the evaluation of teachers. To reiterate: (1) faculty, administration and governing body commitment to the evaluation process; (2) the evaluation program must be an integral part of the total program of teacher education; (3) the evaluation program must be based on objectives and be specific, yet cover the general education of teachers; (4) the program must be longitudinal and continuous; (5) the program must reflect the state-of-the-art of evaluation; (6) the program must reflect thorough planning; (7) adequate computer resources must be available; (8) there must be an effective feedback system; (9) results of the evaluation must be usable in decision making; and (10) the evaluation program must have a self-evaluation component. If these characteristics are not inherent in a program of teacher evaluation, then the system may not be adequate, and consequently the results may not be usable for program development and improvement.

Although this paper describes evaluation of classroom teachers, grades K-12, the design characteristics may be applied to almost any model for the evaluation of personnel completing programs oriented toward professions in the schools (e.g., principals, superintendents, curriculum supervisors, counselors, and school psychologists). It is recommended that any institution establishing a program for teacher education evaluation give serious consideration to the design characteristics outlined in this paper. It is felt that the institution will have fewer problems, a more efficient system that will meet NCATE standards, and furthermore will be more useful for program development and improvement.

References

Standards for the accreditation of teacher education. Washington: National Council for the Accreditation of Teacher Education, 1979.

Hord, S. M. & Hall, G. E. Teacher education program follow-up studies: A collection of current efforts. Austin: Research and Development Center Teacher Education, The University of Texas, 1978.

FINANCE AND RESOURCE ALLOCATIONS RELATIVE TO TEACHER EDUCATION EVALUATION

Edell M. Hearn
Tennessee Technological University

It is sad but true that many, if not most, institutions of higher learning have neglected the evaluation of their graduates. Teacher education in this instance, however, has gone far beyond most of the other professions. Central administrations at most colleges and universities have been reluctant to fund the programs for the evaluation of graduates. This, indeed, is strange in a time when more and more emphasis is being placed on the "quality" of the instruction being provided; namely, how well the graduates perform. This impetus without a doubt has been related to the expenditures for various programs.

In Tennessee the Tennessee Higher Education Commission has implemented a system whereby each institution is evaluated annually based upon several criteria that supposedly relate to the quality of instruction. One important criterion is how well graduates perform on specified objective measures. This is one of the first efforts in the nation to conduct such an evaluation. Irrespective of the various constraints in conducting evaluations of graduates, the fact remains that if programs of teacher education are to be improved there must be an ongoing evaluation of how well the graduates perform as they exit their programs of study and after they enter their professional roles. No matter how critical one may be of teacher education, it is obvious that the profession is trying hard to increase program effectiveness based upon results from follow-up studies and evaluation of graduates. The only national accrediting association for teacher education (NCATE) is placing more and more emphasis on the evaluation of teacher education programs and graduates. In 1979 more institutions were denied accreditation because of deficiencies in governance and evaluation (Standards 1 and 6) than in any other area (NCATE, 1980).

The entire issue of denying institutions NCATE accreditation if they failed to meet all of the standards in a given family of standards, namely, the 1s, 2s, 3s, etc. which would result in automatic denial of accreditation, came about because so many institutions were failing to meet the standards on evaluation and governance. The first vote of the Council on this issue failed by one vote. A year later the issue was again raised by the Standards Committee and the Council voted to approve by two votes. Since then, the Council has asked and approved holding in abeyance any further action until more input can be gathered and analyzed from the various constituencies of the Council. Stemming from the actions of the Council related to the family of standards issue was the development of Evidence Questions relating to all of the standards and a modification of the standards on governance.

It goes without saying that teacher education evaluation must be considered a necessary part of overall teacher education programs. The importance of teacher education evaluation is unquestioned. This being true, irrespective of the NCATE or other standards, there must be adequate resource allocation and funding for the many activities that must occur. Those individuals who control resources in an institution must foresee that evaluation is absolutely necessary for program modification and improvement. If a program is to be accredited or reaccredited by NCATE, there must be a systematic effort to evaluate "program and product" in teacher education. This evaluation must be directly related to the objectives of the program.

The cost of effective and meaningful teacher education evaluation often requires a re-examination of available resources within the control of the institution. It is unfortunate, but it does not seem likely, that the prospects for outside funding for evaluation will be forthcoming. Very few funds from outside sources have been allotted to evaluation efforts. In some institutions that have education foundations, some funds have been utilized for follow-up studies in the evaluation of graduates. In the main, however, it appears that the finance and resource allocations will have to come from the university, college, or department. The amount of finance and resource allocations is dependent upon the comprehensiveness of the teacher education evaluation system. Probably, the greatest cost will occur for the initial implementation of the evaluation system. After the system has become operable, the maintenance of the system will not be so costly. During the first year there must be considerable resources applied for staff development time, for planning, training, etc. The key factors related to cost, of course, will be based upon the design, instrumentation utilized, and implementation of the system. The amount of money to be expended will vary also according to the size of the institution, the number of graduates, sampling techniques, whether direct observations of graduates are to be made, etc.

As a medium sized, multi-purpose university, Tennessee Technological University has developed an evaluation system based upon the AACTE model. Basically, the system incorporates four modes of gathering evaluative data: (1) students; (2) self; (3) immediate supervisor, principal, etc.; and (4) trained observers. More information concerning this system is available from the College of Education, Tennessee Technological University, Cookeville, TN 38501. Following is a summary of the approximate project costs at Tennessee Technological University for 1979-80 (Ayers, 1976).

Estimated Yearly Costs for Conducting Follow-up Evaluation Studies at Tennessee Technological University

3, 1/2-time Graduate Assistants for 9 mos. academic year at $2,385 each	$7,155
1/4-man year for Secretary at $9,400/yr.	2,350
Mileage for visitation in classrooms 7000 miles at $0.20/mile	1,400
Communications (postage and telephone)	600
Supplies and Printing (paper, envelopes, forms, etc.)	500
Total Direct Costs	$12,005
1/4-man year for Professional Employee	6,000
Indirect Costs (overhead)	3,000
Total for all Costs	$21,005

Elimination of the requirement for extensive visitation could potentially reduce the Direct Costs for the project to about $5,500 and the Total for all Costs to about $11,000.

Reasonable follow-up studies can be implemented on a limited budget and still provide very useful information. Follow-up mail surveys can be carried out at a cost of about $0.50 per subject (exclusive of personnel costs). Such surveys are essential for gathering routine information for the development and implementation of larger studies such as the Tennessee Technological University Teacher Evaluation Model.

The Evaluation Model can be implemented at smaller institutions with limited budgets. As was noted earlier, data about the graduates are collected from four sources, i.e., the graduate, the supervisor or principal, students of the graduate, and independent observers. The largest costs are associated with sending independent observers to work in the schools to gather data on a firsthand basis. If this later step can be eliminated, the costs associated with the project can be greatly reduced. At the present time studies are underway to predict the results of the data gathered by independent observers based on data from the other sources.

Personal data about graduates can be gathered by mail surveys, from university records, and telephone calls. Information from employers can be gathered through structured, mailed questionnaires and telephone calls. Information from students of the graduates can be gathered by mailing survey instruments to teachers and in turn asking them to administer these to their students. Direct evidence of classroom interaction can be collected by such individuals as student teaching supervisors or principals (It should be noted that these individuals must be carefully trained). Also, data can be collected by use of audio tapes or video taping in each teacher's classroom. Again, these items can be handled by use of the mail system.

Although the validity and reliability of the data may be lower, it is felt that much useful research data can be gathered, using the above techniques, that will aid in satisfying NCATE standards. In turn, and most important, it is felt that improved systems for the training of teachers will result.

One possible way to improve teacher education evaluation systems and decrease costs would be in establishing a national network of institutions seriously involved in teacher education evaluation. Through such a network there could be a sharing of results of particular sources of data or effectiveness of changes made in programs. Also, assistance could be provided in selecting alternatives which would have a higher potential for success. If such a network were established it would be extremely important for all involved to be participating and not simply being involved just to gain information.

In conclusion, following is a list of considerations on how program evaluations may occur by taking advantage of existing resources (Adams, 1981).

1. Take advantage of field-based faculty activities. Supervisors of student teachers and inservice activities provide opportunities for faculty to collect follow-up evaluation data. Appointments may be made with graduates to obtain evaluative information to coincide with other business in a given location or area.

2. When students are evaluated during the preservice program, these data may be designed to provide useful information about the student's progress. Similarly, systematic and standardized supervision of student teachers may provide valuable "exit" data at little or no extra cost. Entry data that can be used for evaluation purposes can be obtained through the inclusion of selected instruments in the initial packet completed by prospective teacher education candidates.

3. Limit the data collection activities to those students in counties or school systems within a reasonable geographic location to the university. This may, however, affect the validity of some of the findings. Care must be taken in studying the data. If the validity is questioned it may require a wider sampling.

4. One approach might be to select bench mark school districts as units for the evaluation and visit all graduates within a school district on a recurring basis. This is particularly viable if school districts will help in collecting some types of data not available through a more limited contact.

5. A major expense category will probably be the data processing required for a complex evaluation system. This is in the form of both the hardware (computer), software (computer programs), and the personnel required to maintain the data set. This could amount to a sizeable cost, particularly for institutions that do not have the technology readily available. One ap-

proach to reducing these costs is through cost sharing, where two or more institutions could share the cost of data processing, storage, and analysis. The specifics would have to be agreed upon, but the costs could be greatly reduced if these institutions shared the costs instead of each one maintaining them separately.

6. It may be advisable to form multi-institutional consortia. This may provide a way to reduce costs, particularly for some aspects of the evaluation system. The problems with multi-institutional arrangements are obvious: institutional integrity, decision making, logistics, competition for students, etc. The one problem that is a disadvantage is the possible reduction in faculty ownership and ultimately rejection of the evaluation process. If this barrier could be obviated, the multi-institutional approach might be quite attractive, particularly to small institutions.

It is recognized that this paper has not included very many specifics which would relate to all institutions relative to finance and resource allocations for teacher education evaluation systems. Hopefully, however, the case has been made that to improve teacher education programs and products, there must be substantial expenditure of funds allotted to teacher education to carry out a system of evaluation that will be successful. We in teacher education must continue to try to improve our programs. The times call for "proof" that we are truly preparing individuals who can perform effectively in the world of the teaching profession. This cannot be done without a good evaluation system.

References

Adams, R. D. Personal communication, January 30, 1981.

Ayers, J. B. Follow-up studies at Tennessee Technological University. Proceedings of the sixth annual teacher education conference of the southern regional education board-forces for molding management and control of teacher education, 1976, 7-13. (ERIC Document Reproduction Service No. ED 152 695).

WHAT IS THE FUTURE OF TEACHER EDUCATION PROGRAM EVALUATION?[1]

Gene E. Hall
Research and Development Center for Teacher Education
The University of Texas at Austin

Exploring the future of teacher education program evaluation is a difficult and risky task for many reasons. The teacher education program evaluation and follow-up (TEPFU) movement has gradually developed over the last ten years, the principal driving forces having been the work of ten to twelve dedicated individuals and the requirement of NCATE standards 6.1 and 6.2. There does not appear to be a major ground swell of interest in TEPFU just around the corner. Doing TEPFU studies is hard, complex work, sometimes discouraging to prospective researchers, and so not likely to attract large crowds. Study designs are imperfect and the data are dirty. Thus it is unlikely that earth-shattering findings will emerge in the near future which would grab the attention of the mass media. Also, it is a fact of life in an Einsteinian universe that one cannot see the future.

Regardless of the problems, conditions and reasons why the future can't be predicted with accuracy, I accept the challenge to use this opportunity to reflect on the TEPFU movement. I will also comment on study designs and possible influences that I predict for the next few years. In preparing this paper some concrete recommendations manifested, which I would like to offer about the directions I feel the movement should take. These recommendations are within the realm of what we can control. I will also do some speculating about the future in areas where we have little or no control. My recommendations for the controllable future make some sense; for the uncontrollable future my guesses may help us to anticipate some of the pressures that will be placed on us. In neither case do I see profound breakthroughs, great public acclaim or a large number of grants coming to or out of TEPFU activities. I do believe, however, that there are some things we can do to enhance the quality and quantity of TEPFU studies and the use of study findings.

---

[1] The research described herein was conducted under contract with the National Institute of Education. The opinions expressed are those of the author and do not necessarily reflect the position or policy of the National Institute of Education. No endorsement by the National Institute of Education should be inferred.

The Development of TEPFU: 1970-80

Teacher education program evlauation and follow-up studies came into their own in the 1970's. The AACTE stimulated paper by Sandefur (1970), the CBTE movement and the federal emphasis on evaluation of everything catalyzed different ways of thinking about teacher education program evaluation. In the past, program evaluation consisted of course grades, data from the placement office about the number of graduates who went into teaching and the occasional mailed questionnaire to post-graduates to find out what they remembered about their preservice program. In the 1970's the evaluation questions became a little more focused and there was increased emphasis on follow-up studies, and within-program evaluation to a lesser extent. The two research questions most frequently addressed were: (1) What are graduates of preservice teacher education (PTE) programs like after they go into the field? And, (2) What are the immediate effects of different courses and experiences?

With the early lead of Adams and Sandefur at Western Kentucky University and Ayers and Hearn at Tennessee Technological University, data began to be collected about graduates one, three and five years after they left the program. The CBTE movement encouraged program evaluation efforts which looked at the immediate effects of certain program components at such institutions as the University of Nebraska, University of Houston, Oregon College of Education and Weber State College. And in the late 1970's Kevin Ryan recruited six doctoral students to do an in-depth ethnographic study of first year teachers. In addition, NCATE and a few states (e.g., Ohio) began to push for documentation of teacher education program effectiveness and follow-up of graduates.

There were other studies, institutions and persons who could be identified, but for our purposes here the ones cited provide an overall picture of how the 70's developed. One additional trend out of the 70's that is important to note is the gradual development of an invisible college or network of the actors who are concerned with the conduct of TEPFU studies (Cooper, 1979). The Texas R&D Center has participated in this. Note that to date this network has served primarily as an internal support system for TEPFU researchers and not as much for external dissemination and support building.

In summary, during the 1970-80 period the following occurred:

1) TEPFU emerged as a distinct area of study.
2) All studies were done in relation to preservice programs.
3) Studies focused on within-program effects and follow-up of graduates.
4) The number of studies and institutions involved increased.
5) TEPFU became required for PTE program accreditation.
6) Individuals and institutions involved established a network.
7) Some of the same measures were used across studies.
8) The overall amount of activity, number of actors, and data base increased dramatically.

Interestingly, in a decade of high federal involvement in all aspects of education, to the best of my knowledge no TEPFU activity had direct federal support. The awakening of the need for TEPFU studies and the support for them has been done by individual institutions and dedicated individuals with little or no outside encouragement and reward. The TEPFU'ers of the 1970's should be applauded for demonstrating the kind of initiative and commitment that means professional teacher education.

One other characteristic of TEPFU studies in the 1970's that should be mentioned was the "whacking" at different parts of the teacher education continuum without maintaining a picture of the totality. Some institutions and individual researchers devoted nearly all their energies to collecting data in longitudinal studies of graduates. Others carefully tested individual courses and modules, while some focused on initial teacher selection. The studies tended to reflect an overemphasis upon particular points along the teacher education continuum and particular components of a teacher education program, without paying equal attention to the interrelationships with other parts of a program, or the overall continuum. There was, furthermore, very little success in developing effective strategies for using the results of TEPFU studies to change preservice programs.

If the above are a few of the highlights of the 70's, then what would be reasonable recommendations and predictions for the 1980's? These are the topics to be discussed in the remainder of this paper. First, the framework of the teacher education continuum will be described. The continuum will then be used to raise questions about present study designs and to make recommendations about how TEPFU studies should be focused in the future. This we can control. The final parts of the paper will be reserved for a few speculations about what lies ahead, over which we have little or no control. Note that most of my remarks will come from the preservice perspective since that is where the activity has been.

The Teacher Education Continuum

One of the important philosophical pushes at this time is the shift away from thinking of teacher education as a dichotomy consisting of preservice in one cell and inservice in another. Instead of viewing teacher education from the point of view of distinct college courses and random inservice training experiences, teacher education is now being viewed as a continuum of profession-long development. From the teacher's perspective, teacher education is a continuum of experiences from early entry, to preservice teacher education (PTE) to "induction" (i.e., early inservice) to the career-long inservice phase.

Viewing teacher education along a continuum is not only important for program development, it is also important to the design of PTE program evaluation. To illustrate the relationship between the teacher education continuum and TEPFU consider Figure 1. In this figure the various phases along the teacher education continuum are named. The figure also identifies key points for data collection. O's, representing "observations," are used to mark these points. Note that the type of program (e.g. elementary, secondary, CBTE) or the types of measures (e.g., questionnaire, observation, interviews) do not need to be differentiated here. The assumptions are that the framework is simple enough to apply to most teacher education programs and that multiple measures would be required at each O. Although this framework looks overly simplistic, it can be used to point out some of the futility of the "whacking" that was done in the 70's and to make my recommendations for the future.

In designing a TEPFU study or any other study the first step must be deciding on what questions to ask. The research question/s sets parameters on the study design and the variables to be assessed. Some of the different TEPFU questions that could be asked are summarized in Figure 2.

The variables to be assessed obviously dictate the measurement options. For example question 4, "What are graduates of our PTE program like?" would require a different study design than would be required for the question, "How long do teachers maintain the skills that they had at the time of PTE graduation?" The first question only requires one time data collection at point O_3. The second question requires longitudinal data collection at times O_3, O_4 and on to O_t, with the same variables assessed each time. All of these questions and some others could be asked, but all do not seem equally reasonable. Unfortunately, many TEPFU studies in the past have not fully developed their questions or the reasoning behind their designs, as will be illustrated below.

Recommendations for Future Study Questions

Several TEPFU studies in the 70's placed heavy emphasis on the follow-up of graduates. The standard design was to collect data of 1st, 3rd and 5th year teachers. I question the utility and validity of doing more of these studies. Ayers (1980) and Adams (1979) have reported little change in findings across cohorts and that the third year data reflected the "highest level of functioning" (Ayers, p. 29). I contend that it is highly unlikely that the effects of a preservice program can be distinguished from other intervening variables after the first year of inservice. The induction phase, which is not addressed in the U.S. but is in the U.K., is such a powerful treatment that any residual unique preservice program effects would be washed out. Even if they were there, it would be impossible to measure them with today's designs and measurement technologies.

If some agency has need for data about teachers' skills as they move through their career, then that agency should support the necessary longitudinal study. It is not directly useful to institutions who are responsible for initial training. The TEPFU focus for preservice programs should be on (1) assessing teacher knowledge (i.e., cognitive competencies) and (2) their teaching skills (i.e., performance competencies) at the time they exit the program, not long term follow-up. Some useful program development information might be obtained from following a sample of teachers during the induction phase, but there is no apparent reason for the preservice institution to follow them into full inservice. Now, if a teacher education institution plans to become involved in teacher education for the induction phase (TEI) then further follow-up data would be warranted.

Another point, which has recently been proposed by deVoss and Hawk (1980) and Borich (1980), is collecting teacher "need" data from inservice teachers as a part of PTE/TEPFU activities. Now, I have always had a hard time understanding the logic behind needs assessments. I don't see how a teacher can tell you what they need when, by definition, they don't know what they don't know. So how can they tell you what they need? They can identify problems and concerns though.

But the concerns model (Fuller, 1969) and other developmental models demonstrate that teachers at different points in time perceive that they have different kinds of problems. If the developmental models are correct, a teacher changes in concerns as s/he moves along the teacher education continuum. Of what use to a preservice program is extensive information about the needs of teachers five years after graduation who are significantly more mature in their professional development? Right now we don't even address first-year teacher concerns in preservice programs. Once again, I would argue against multi-year follow-up studies for preservice program evaluation.

As Tom (1980) has pointed out, PTE institutions now receive NCATE credit for doing TEPFU studies. Neither the quality of the study nor the quality of the PTE graduate affects program certification. I suspect that quality checks on graduates will continue to be pushed by state competency tests. I hope that these quality checks will consider both cognitive and performance data. Clearly, TEPFU studies need to do this. Survey questionnaires with course happiness coefficients are very weak data to use in refining programs. Hopefully teacher training institutions will establish their own quality controls without outside force being applied.

The movement toward "common core" instruments is encouraging (Schalock, 1979). The only way that cross program comparisons will be possible is if the same data are collected for each program. The resistance to agreeing on measures may indicate some resistance to having cross program comparisons made.

Another step which needs to be taken is to collect the same data at each point along the continuum. Some changes in measures should be done to incorporate new methodologies, but any study questions that require comparison at two points in time require that the same data be collected at both points. Although this is an obvious point, why are there so few instances of the same variables being measured for preservice and then in follow-up studies?

Future TEPFU studies should look more closely at within-PTE program features (e.g. courses, faculty, modules, field experiences) to determine what effects each has. The global assessment of teacher characteristics at O_3 does not help in program refinement unless we also know what the specific effects are of different parts of the program.

In summary, my recommendations for the near future in areas which we can control are:

1) Cut follow-up studies back to first year follow-up only.
2) Focus studies on assessing the effects of basic components of the PTE program.
3) Continue the move toward common core measures.
4) Give some thought to what the significance is of the induction phase.
5) Be sure the study design will address the prespecified study questions.
6) There should be study questions.
7) Collect the same data at different points in time.
8) Although not mentioned specifically above, limit the studies to what can realistically be done, and insure that the findings are used.

The Uncontrollable Future

Much of the future is unpredictable and much cannot be controlled. My guesses of the moment about TEPFU over the next several years are as follows:

1) Continued boot strapping. I do not see policy makers, funding agencies or program developers becoming overly excited or even particularly in-

terested in TEPFU findings. I do not wish to be a prophet of doom and gloom, but I just do not see a pot of gold, glory, or even much interest in the work of a few very dedicated individual efforts.

2) New study sophistication. The "how to's" have been worked out and the track record of the 70's demonstrates that TEPFU studies are manageable. We also have baseline data on many programs. In the 1980's the studies should be more focused, more sophisticated in design and should use measures with increased reliabilities and validities.

3) Network building. The network of persons and institutions involved in TEPFU activities should continue to grow. The network can provide mutual support and increase collaboration. It is important to the movement that all members be supportive. Even when there are ideological differences, it will be important to address these in positive ways. The network is too fragile to support warring camps. Criticism in the form of attacks, no matter how well intentioned, will fracture rather than help to build a stronger network.

4) Micro-computer data bases. With increasing use of micro-computers for program management and monitoring, an expanded data base should become readily available for TEPFU studies. For example, many of the process and cost questions should be addressable by analyzing program management data.

5) Use of TEPFU findings. Developing strategies that will facilitate use of TEPFU findings to modify programs will continue to be a problem. We are a data collection profession, not a data using one. The problem is much larger than TEPFU, and one that seems to defy solution. Perhaps this very challenge is one that TEPFU people could address as part of their research.

6) Inservice TEPFU. Teacher education occurs across the professional continuum. In the 1980's induction and inservice teacher education program evaluation and follow-up studies should emerge as important activities. There is little or no data on the effects of induction and inservice programs in this country. In fact the emphasis of the TEPFU movement could shift toward these activities. One spin-off of this could be a larger number of individuals and institutions who are interested in TEPFU activities and this could lead to more national attention and better funding.

The preceding have been some thoughts about the future of TEPFU. There are many teacher educators who see the importance of doing teacher education program evaluation and follow-up studies. There are even some who will use the findings. The movement was born in the 1970's; now all we have to do is help it grow and contribute to the improvement of teacher education at all points along the continuum.

References

Adams, R. D. Critical findings from the teacher preparation evaluation program: Teacher entry, retention, and change. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1979.

Ayers, J. B. A longitudinal study of teachers. Paper presented at the annual meeting of the American Educational Research Association, Boston, April, 1980.

Borich, G. D. A needs assessment model for conducting follow-up studies. Journal of Teacher Education, 1980, 31(3), 39-42.

Cooper, J. M. Improving teacher education program evaluation. In G. Hall, S. Hord, & G. Brown (Eds.), Exploring issues in teacher education: Questions for future research. Austin: Research and Development Center for Teacher Education, The University of Texas, 1979.

deVoss, G. & Hawk, D. Follow-up models in teacher education. Submitted for publication to Journal of Evaluation and Policy Analysis, 1980.

Fuller, F. F. Concerns of teachers: A developmental conceptualization. American Educational Research Journal, 1969, 6(2), 207-226.

Sandefur, J. T. An illustrated model for the evaluation of teacher education graduates. Washington, DC: American Association of Colleges for Teacher Education, 1970.

Schalock, H. D. Follow-up studies on graduates of teacher preparation programs in Oregon. Paper prepared for conference on follow-up studies in teacher education sponsored by the Ohio Department of Education, Columbus, OH, November, 1979.

Tom, A. R. Chopping NCATE standards down to size. Journal of Teacher Education, 1980, 31(6), 25-30..

Figure 1: The Teacher Education Continuum

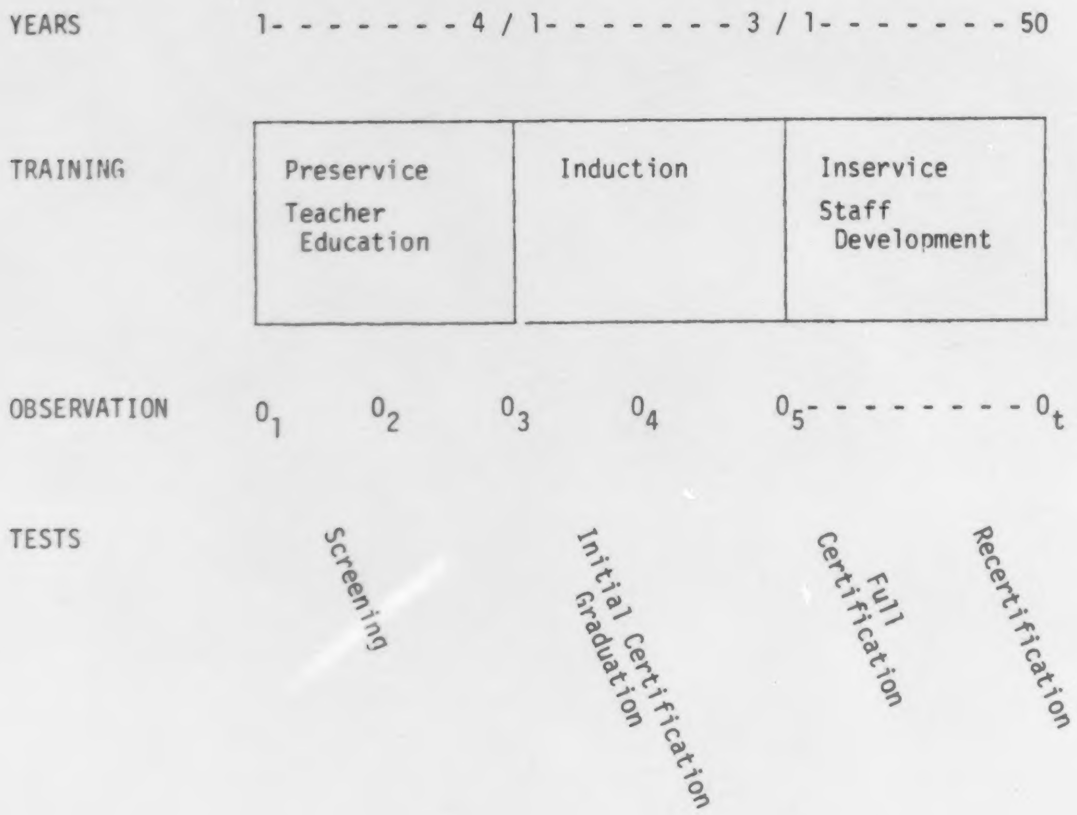
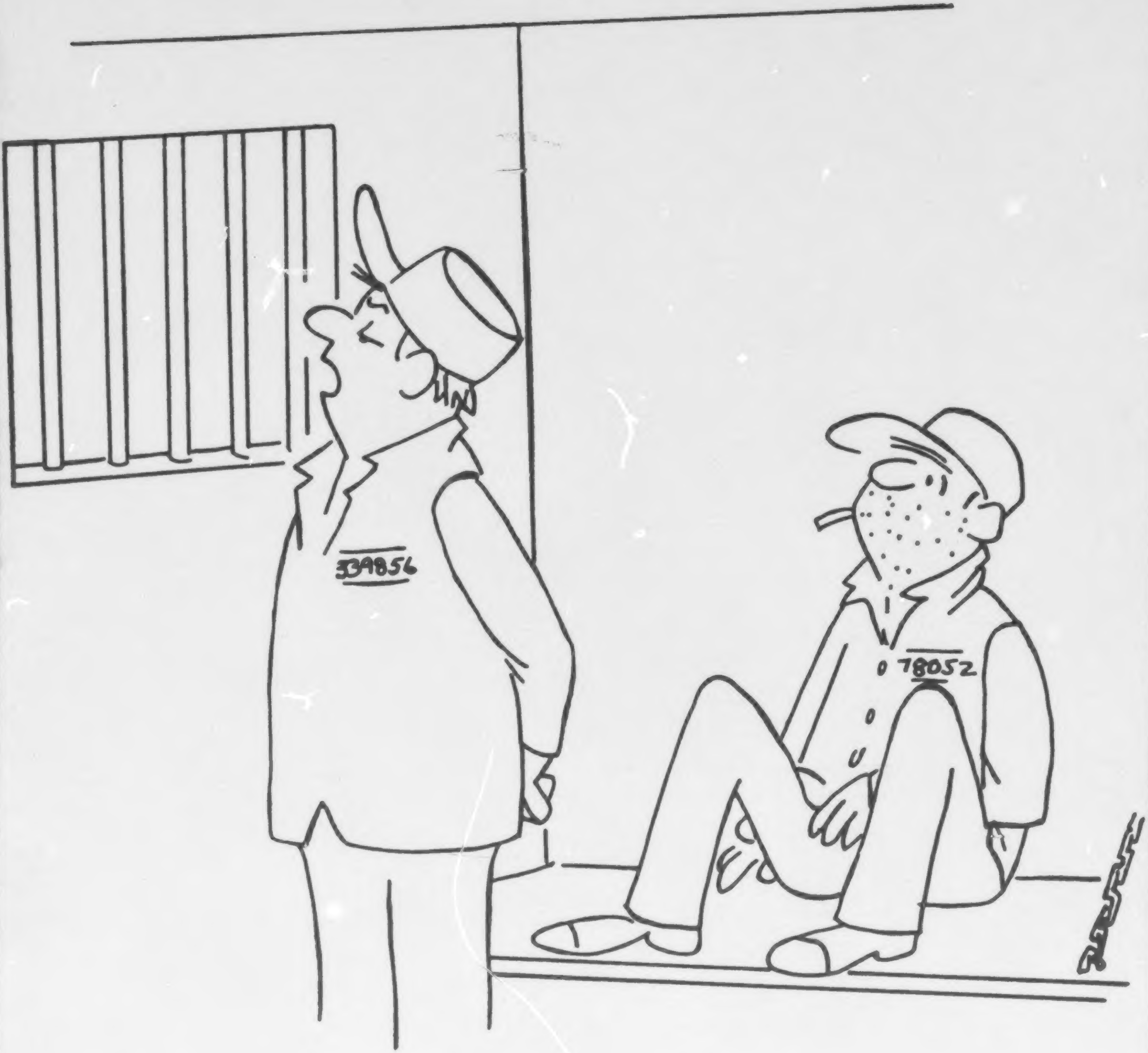

YEARS
1- - - - - - - 4 / 1- - - - - - - 3 / 1- - - - - - - 50
TRAINING
Preservice
Teacher Education
Induction
Inservice
Staff Development
OBSERVATION
0_1
0_2
0_3
0_4
0_5- - - - - - - - 0_t
TESTS
Screening
Initial Certification
Graduation
Full Certification
Recertification

Figure 2: TEPFU Study Questions

Number	Data	Questions
1	0_1	What are the characteristics of preservice teacher education applicants?
2	0_2	What are the contents and processes used in PTE?
3	0_2	What are the immediate effects of particular PTE processes and contents?
4	0_3	What are the characteristics of PTE graduates?
5	$0_3 - 0_1$	What are the immediate effects of the PTE program?
6	$0_1 - 0_3$	What is the relationship between characteristics of graduates and their entry characteristics?
7	$\frac{0_3 - 0_1}{0_3^{1} - 0_3^{1}}$	How do the graduates of one program compare with the graduates of another program?
8	0_5	What are the characteristics of teachers at the time of their full certification?
9	$0_5 - 0_3$	How does the induction phase affect teacher characteristics?
10	$0_t - 0_5$	How do teacher characteristics change across years of inservice?
11	$0_8 - 0_7$	What are the effects of a particular inservice?


339856
78052

". . . And then someone on the accreditation team said, 'You don't meet our program evaluation and follow-up standard'!"

TEACHER EDUCATION PROGRAM EVALUATION
A SEARCH FOR ALTERNATIVES

Tom Savage
Texas A&M University

Any discussion of teacher education program evaluation and follow-up is certain to raise the anxiety level of teacher educators. The discovery that weakness in program evaluation is a major cause of accreditation denials by NCATE is certain to elevate blood pressure of those charged with program evaluation responsibilities. What causes the anxiety is a lack of knowledge of alternative evaluation models. This void is clearly illustrated by the emphasis that has been placed on the practice of mailing questionnaires to graduates and their principals. When the obvious weaknesses of this data gathering practice are pointed out to teacher educators, they have difficulty stating alternatives which are feasible and cost effective.

Perhaps the first step that is needed in generating alternative models is to address some basic questions and assumptions which guide current practice. This is an especially crucial task because many teacher educators feel that accrediting agencies are operating on some assumptions that have questionable validity. A case in point is the assumption that data supplied by graduates and principals of graduates is a sufficient data base for making program decisions.

The typical expectation that a follow-up of graduates is needed to determine program effectiveness appears to have a certain amount of face validity. However, a bit of probing uncovers some conceptual weaknesses. This requirement presumes that there is a great deal of similarity in teaching environments and a consensus of criteria used to evaluate teacher effectiveness. Such is simply not the case. The criteria used by one principal might be quite different than the criteria used by another principal. This expectation seems to assume that all data are of equal worth and only one interpretation of the data is thought to be feasible, regardless of the criteria used by the person performing the evaluation or the environment where the teacher is employed. If a negative evaluation is received it is interpreted to mean that something is wrong with the preparation of the individual.

Proponents of the follow-up studies use the industrial model as an example. They point out that just as industry is willing to guarantee a product, so should teacher education institutions be evaluated by the performance of their products. It should be noted, however, that most manufacturers guarantee a product only if it is used as intended. A case might be made that many teacher education graduates are placed in settings where they are unable to perform as intended.

Another unstated assumption of follow-up studies is that the teacher preparation program has some long term effect on a given graduate. There are grounds for challenging this assumption. Numerous intervening variables such as the nature of the setting, pressure from teaching peers, the style of the administrator and the availability of inservice opportunities either reinforce or negate the effects of a teacher preparation program. To expect a young inexperienced teacher to persist in certain behavior patterns acquired in their preparation program that might be at odds with the status quo is unrealistic.

While most teacher educators do accept the notion of being held accountable for the quality of their graduates, they wonder if they should be held accountable for the performance of graduates in all types of situations and for an indefinite period of time following graduation. While there are many calls for "proof" that teacher education institutions are preparing individuals who can perform well as teachers, what sort of data is needed to supply that proof? This is one of the major issues which needs to be addressed.

The Hall paper in this publication assists in the search for alternative models by suggesting that teacher education be viewed as a continuum which includes several evaluation points. Each of these program evaluation points poses a different question and requires different data. It might be added that different agencies should be responsible for data gathering at these different evaluation points.

For example, the program evaluation conducted by institutions of higher education should be focused on the program objectives and the accumulation of evidence relating to whether or not a candidate has met those objectives. This type of program evaluation model would have the advantage of providing feedback that could be tied to specific program components. This would facilitate program improvement and change. A rather rigorous set of exit level requirements could include such things as the impact of the student teachers on learners' achievement. This would require greater emphasis and more attention to the student teaching component. But this is something the institution can and should control. Any discussion of program evaluation by accreditation agencies and other interested parties could then be directed to the appropriateness of the program objectives and the quality of evidence used to support program and certification decisions.

The follow-up of graduates might best be assumed by the state education agency. The model being developed in Georgia which includes regional assessment centers staffed by individuals specifically trained to do in-class observation and charged with the responsibility of evaluating all first year teachers fits well with this concept. Data could then be gathered regarding the performance of graduates on a statewide basis using a consistent set of criteria. This would be a much sounder evaluation process than what is currently practiced.

In conclusion, what is needed in order to advance the state of the art are alternative program evaluation models. The absence of viable models has led to high levels of anxiety and frustration. A critical element in generating these alternatives is the questioning of some basic assumptions. When this is done, alternate models can be proposed which meet the criteria of an effective evaluation program.

Participants

Raymond L. Archer
Northeastern Oklahoma State University

Caseel D. Burke
Weber State College

Rodney H. Johnson
University of Wisconsin -
Eau Clair

Bill Lampkin
Baylor University

PROGRAM EVALUATION: WHAT, WHEN, AND HOW?

Ted White
Glassboro State College

There appear to be four general areas in which discussion was focused by the participants in our group. The areas break out something like this: 1) What should we be evaluating?; 2) When should we evaluate it?; 3) How should we evaluate it?; and 4) What should we do with it? Each question is discussed separately.

What should we be evaluating?

In the case of this first question, there was consensus that the overall effort at program evaluation should determine the extent of match between program outcomes and professional community needs. That is, an effort needs to be made to shape an evaluation system that will yield data useful in assessing the extent to which this match exists or doesn't exist. Of course, the position presumes that, on the one hand, we have a handle on professional community needs and on the other hand we can devise an evaluative system capable of assessing the degree of program/community match. The second part of the first question (what) involved the relationship between the process evaluation and product evaluation of the program. Follow-up by itself does not represent total program evaluation in the same way that front-end evaluation by itself does not constitute total program evaluation. There is a need to articulate internal instructional evaluation (courses, admissions, instruction, clinical work) along the way with the evaluation methods to be employed during follow-up.

When should we evaluate?

I have to an extent touched upon the second area of discussion, namely when should we evaluate programs. Again, there was a general consensus that follow-up evaluation by itself will not speak to the multiple concerns in program evaluation. A greater effort needs to be undertaken in longitudinal data collection with prescribed access points for data collection identified.

The opportunity will then be available to use the data formatively with a bit more precision than were we to rely solely on follow-up data exclusively.

How should we evaluate?

The third general question and the one which generated the most discussion was: How should we evaluate what we're evaluating? There appeared to be widespread agreement that an early identification of instructional and program objectives would be a necessary precondition for the design and implementation of comprehensive evaluation. The full involvement of the teacher education community (perhaps including the practicing profession) would be essential for building ownership and giving the evaluation process validity. Third, overt teacher behavior during direct classroom observation is insufficient by itself to be considered program evaluation. What and how teachers think may be as important as how they act. Also, long-term follow-up (beyond one or two years) may not be very reliable and therefore probably not cost effective. That may be true to the extent the training program is conceived as a four or five year inclusive period with no systematic continuing involvement of the program beyond the fourth or fifth year. Should the system of training be viewed differently, that is occurring beyond a fixed four or five year period, the evaluation strategy will likewise need to be viewed differently.

What should we do with it?

The last question dealt with in this group was the "so what" question: What should be do with it once we've done it? Again, there was general agreement on two points. First, there needs to be a deliberate effort to create the feedback loop between evaluation and program development (or redevelopment as the case may be). If we complete the evaluation only to have the data spend its days in archives we obviously are not serving the profession's best interests.

Lastly and of substantial importance was the universally held belief that TEPFU should provide the vehicle for the presentation of evaluation programs currently in place or under development with a discussion of conceptual features, operational preconditions, implementation problems and benefits, cost effectiveness and most importantly what difference has been made as a consequence of the program. The sense of sharing that was evident in this entire program needs to be expanded in a network of people ready to share ideas, success and failure. For us to do less would be to ignore the evaluation evidence that the session produced.

Participants

Kenneth Elsner
Central State University

Gary R. Galluzzo
Glassboro State College

Robert D. Krey
University of Wisconsin-Superior

Juana A. Mendes
University of Puerto Rico

William A. Person
Mississippi State University

PROGRESS IN EVALUATION OF TEACHER EDUCATION

Vernon Dravland
University of Lethbridge

Evaluators and researchers in education operate within a great dilemma. Aspects of this predicament were evident at this conference. To illustrate, let us draw an analogy between education and one of the sciences, for example, mathematics. The knowledge bases of education and mathematics are, perhaps, too diverse for comparison. However, we can make a rather interesting comparison if we compare the levels of ignorance in the two areas. The ignorance in education is very unsophisticated, while that in mathematics is very sophisticated. It takes another mathematician to recognize the ignorance displayed by a mathematician. However, even the "most ignorant" can propose what may appear to be very good solutions to very complicated problems in education. The key word here is "appear" since research in education is actually much more complex than that in mathematics, partly due to the large numbers of extraneous variables which are difficult to control. To use another example, during the development of the atomic bomb in the 1940's, scientists were able to define areas of ignorance that had to be overcome in order to develop the device. Teams of researchers were then assigned to each problem area and one by one the areas of ignorance were eliminated. Defining areas of ignorance in science is not always possible, but more likely in science than education.

We do not know what we do not know in education, and therefore cannot as easily define the variables or areas that require study before we can solve a particular problem. Research in both education and science tends to create more questions than answers. However, in education these questions tend to diverge from the initial problem. In the physical sciences if an answer is not found, the research tends to at least help clarify the question. This does not imply that there are no surprises in the results of research in the sciences, but the questions that arise are more likely to be convergent in nature. The situation in evaluation is an outgrowth of the situation in education research. Since our knowledge base, which normally comes from research, is inadequate, the theory of evaluation is inadequate.

It became obvious from the paper presentations and discussions at this conference that researchers and evaluators in education are concerned not only with the knowledge base, but also with instrumentation, procedures and administration of projects. Again, in comparison with the instruments available for physical scientists, those available for education researchers are highly unsophisticated. Researchers in the sciences are continually developing new procedures. These procedures and techniques have always been much more precise and universally accepted than those in research or evaluation in education. In the sciences eliminating areas of ignorance is like fitting pieces into a jigsaw puzzle. In education research and evaluation we seem to lack a well-articulated "framework" within which we can define our tasks and our areas of ignorance. This may be the reason for lack of a unified thrust in education.

During this conference the problem of evaluation of teacher education programs was attacked from many and various points of view. Discussions focused on everything from the initial conception of the evaluation design through the administration of the project and the resulting changes in the teacher education program, back to the conception of a new design. In the discussion group it became quite obvious that important philosophical differences among the various institutions might necessitate that different approaches to the evaluation of teacher education programs be developed. Despite the desire expressed by participants for adequate evaluation of all aspects of the teacher education program, the primary concern centered on a system for generating changes in the program based on the results of the evaluation. Therefore the central problem appears to be the establishment within each teacher education institution of an organization which would design an evaluation system that would (1) be acceptable to the people responsible for the teacher education program, and (2) develop a system with the most effective "state-of-the-art" procedures and techniques, in order to maintain credibility within the institution as well as in the teacher education academic community. The presenters made many excellent recommendations that could help solve problems confronting those involved with evaluation of teacher education programs.

Education research and evaluation are very young compared with research in the sciences. However, the last ten years have seen considerable progress. We have a more adequate knowledge base and hence a more adequate theory of evaluation.

Participants

Richard L. Aagness
Idaho State University

Harley Adamson
Weber State College

Randy Elmore
Georgia Southern College

B. W. Walker
Slippery Rock State College

GROPING... AND COPING...
WITH
CONTRADITTORY PROGRAM EVALUATION REQUIREMENTS

Gary G. deVoss
The Ohio State Unieersity

Lyn Gubser's opening remarks made the most impact on our study group of any speaker's. His remarks made such an impact, as a matter of fact, that they caused a certain amount of restrained panic among the study group. Several issues were raised concerning theory: "How should follow-up attempt to separate program effects from individual or extra-program effects?" or "What are the hidden complexities within program and teacher-effectiveness evaluation?" But far and away the most frequently-repeated question was, "How can Lyn Gubser tell us not to use questionnaires when we only have money for questionnaires?" After processing similar questions a bit, the underlying question that emerged asked what ways acceptable alternative models of follow-up could be disseminated and implemented and be given positive approval by a visiting site evaluation team from NCATE. This study group was saying it wanted to get away from questionnaires and try new avenues, new techniques, new approaches, but was scared to death to do anything which might be perceived as unsound or unsafe. Ergo the profusion of one-shot questionnaires Adams spoke of--not only does the questionaire provide safety in numbers ("Everyone else is doing it"), but it shows that something is being done, even when no one in the institution really knows what to do or really knows what NCATE might be looking for as the benchmark of a satisfactory program.

Without doubt the most significant points made in the ensuing discussion were that a) with some help, institutions could design creative, meaningful follow-up studies quite easily, b) institutions were poised and ready to implement these innovative strategies as soon as NCATE made it clear that they could be implemented without the ultimate penalty--nonaccreditation, and c) a set of general guidelines would be vastly appreciated as schools experimented with different ways to collect data. The first point was cogently demonstrated as the group spun off myriad ideas for doing follow-up evaluations. These include: modifying the questionnaire technique by reducing the number sent to 10-25% of the population (randomly drawn), then attempting to contact and obtain a return from 90-100% of the sample; interviewing by telephone to reduce transportation costs; asking volunteer groups of teachers to come to the university and sit in panel with faculty; starting the documentation process early in the undergraduate career; enlisting students, teachers, and staff personnel (such as university supervisors) in the task of keeping voluntary "critical incident" logs; using extremely small samples of teachers, but keeping extremely "thick" records of their experiences, concerns or problems; using traditional ethnographic techniques either at the same time, or in place of, survey-type techniques.

These suggestions, some of which are already being tried in institutions around the country, emerged in 20 minutes of discussion; it was truly exciting to see how creative institutional representatives could be when the sky was the limit, not NCATE standards.

The second point needs little reiteration. It may suffice to repeat that "experimental" projects are being piloted in various institutions at this time, and that institutions represented in this study group were ready to give anything a try so long as they would not be reprimanded sometime in the future for their ingenuity now.

The third point contains a clear request for facilitation, such as Hall called for when he spoke of networking. Institutions are not afraid of reinventing the wheel, but do not like to do so because first, it makes them look bad when visited, and two, it is cost-ineffective. Two activities were called for here. One was the prerequisite that NCATE begin to actively encourage experimentation and perhaps even waive the letter of program evaluation law for a period of time if the institution can show that it is pursuing some sort of program evaluation practice. The second activity would be to disseminate a Handbook of Possible Program Evaluation Strategies, which would contain several variations on the Western Kentucky model, and certainly several alternatives to the one-shot questionnaire model. This handbook, perhaps coupled with a toll-free "program evaluation hotline" number, would appear to be the equivalent of an on-site expert. The group felt such a handbook could be written today; they also voiced the opinion that they would certainly try anything if a guarantee against censure for daring to do so would be made by NCATE.

As one participant remarked, "It's time we started meeting institutional needs rather than NCATE needs. The one-shot questionnaire ain't great and it doesn't tell us anything, but it meets NCATE's needs. So we're going to keep sending 'em out until we hear different..." Maybe this session was the start of something different. Let's hope so.

Participants

Dr. Harley Adamson
Weber State College

H. H. Carpenter
Nicholls State University

Mildred Chapman
Taylor University

Catherine Cornbleth
University of Pittsburgh

Willard Korth
University of Pittsburgh

Earnest Dees
Alabama A&M University

FACULTY REACTIONS TO PROGRAM EVALUATION:
A LITANY OF "YES, BUT'S..."

Len Froyen
University of Northern Iowa

Evaluation, although critical to effective program planning and development, does not receive nearly the attention it merits. The two most commonly cited reasons for this situation are 1) the cost and logistics associated with conducting evaluation, and 2) the difficulty of converting the results into program improvements. Faculty resistance, however, is another factor which helps to keep evaluation in a subordinate position.

Faculty prize autonomy, and any effort to improve accountability is viewed as an intrusion on academic freedom. The keystone of effective evaluation--precisely specified and agreed-upon objectives--introduces expectations and prescribes procedures that run counter to the preferred academic life of most professors. It is therefore not surprising that department faculties give only cursory attention to consensus building when they establish program objectives. Even less consideration is given to ways of checking compliance with objectives. As long as program evaluation is not taken seriously, these pro forma sessions satisfy the trappings of academic life. However, when evaluation becomes an instrument for examining practice with an expressed purpose to improve upon it, the agile minds of faculty go to work. The litany of "yes, but's" goes something like this. Evaluation is undoubtedly a good and desirable thing, but...

1) You cannot really measure the truly significant outcomes of instruction; evaluation tools are not sensitive to the things that really count.

2) The costs of time and money to conduct a sound and defensible program are not warranted by the limited use one can make of the results.

3) We already know what needs improving; the results only confirm our worst suspicions, and the changes deemed important are those that are doomed for political and economic reasons.

4) It is impossible to secure the cooperation of those who are to be

evaluated, and when they do cooperate, they elevate their efforts to produce the hoped-for results.

5) We cannot possibly separate the effects of training from those that are due to a host of other factors.

6) Evaluation so often becomes an end in itself, serving those who administer the program rather than those for whom the data are supposedly collected.

7) The most effective and verifiable methods, performance-based appraisals, are used on small samples, and this makes the results highly questionable.

8) The sheer complexity of the criterion variables precludes valid and reliable measurement.

These are compelling arguments. Evaluators are hard-pressed to answer these criticisms.

Faculty have created a line of defense that can only be penetrated by more sophisticated studies or by external pressures which are persistent and unequivocably focused. The state of the art will make it difficult to counter the foregoing criticisms for some years to come. Yet faculty must be made part of the solution. They may only join the effort when legislative mandates, state departments of education, professional education associations, and accrediting agencies demand that faculty participate in projects designed to identify and verify the products of instruction.

In the absence of such pressures to break down faculty resistance, evaluators might devote more time to using already available data and assisting those who design the tools and administer the mechanisms for collecting these data. For example, several states now use the National Teacher Examination or a state licensing examination. The objectives that comprise the test blueprint may be used to establish baseline requirements on the local campus. Predictable dissatisfaction with these objectives can become the first step in bringing faculty into responsible participation in program evaluation. Reports from local school systems, based upon their faculty evaluation programs, can also serve a similarly useful purpose. Performance samples taken from the evaluation instruments of several school systems can serve as an initial method of assessing job-related objectives in the teacher education curriculum.

The questionnaire has long been the standard fare of the program evaluator. Much remains to be learned about the effective use of this survey instrument. Time, cost, and the simple logistics of collecting data will continue to make this approach an appealing one. Refinements in the areas of instrumentation, response rates, reporting procedures, and transformation of results into corrective measures should be encouraged and supported through collaboration among program evaluators. We may be well advised to concentrate our energies upon these activities while seeking ways to create a setting that is more receptive to other data collection and analysis procedures. The time saved by relying on more conventional and tried techniques may be most profitably spent in helping faculty to understand and accept the positive contributions of evaluation.

Participants

LeOra M. Lipe
Louisiana State University

Edwin L. Lyle
Association for Teacher Educators

M. L. Roberts, Jr.
The University of Alabama

E. J. Schacht
Northern Michigan University

THE USE OF PROGRAM EVALUATION RESULTS: PROBLEMS AND SUGGESTIONS

Ann Hungerman
University of Michigan

The topic of our group discussion was identified immediately and with easy concensus as the USE of evaluation results. This topic had been mentioned in various ways in five of the six papers presented in the two hours preceding the discussion. Selection of this topic also revealed that everyone in the group was involved in program evaluation and accepted the fact that the use of results is a political necessity as well as a requirement for effective program development.

Our group first identified several immediate problems for putting evaluation results to use, and then addressed these problems in terms of three questions: To what different uses can or should program evaluation data be put; how can we incorporate program evaluation results into program components; and, how can we guarantee that results will be used? These problems and questions are discussed below.

There were several problems which were recognized by the group to be significant. First of all, different uses require different instruments, procedures or design, not all of which are always compatible. For example, longitudinal comparisons require similar instruments, whereas indicated course revisions may suggest changed instruments. Another problem noted with longitudinal studies is that the validity of testing the effect of the preservice program after the first year out in the field is lessened by the effect of graduate study. Furthermore, those experienced in some form of program evaluation report often being caught in a tradition of very informal use of results, making it difficult to switch to more systematic usage, while those just beginning program evaluation need models and guidelines to assist them. And, not to be ignored is the danger that, during evaluation, we can identify weaknesses much more confidently than we can reach consensus on a single definition of a "good" teacher, which sometimes leads to a tendency to panic at evaluation results, rushing precipitously into course or program change.

As these problems were identified, it became clear that there are no easy solutions, but the group did look at questions which might cover more generally the complications involved in results usage. The first most likely question to ask is, to what different uses can or should program evaluation data be put? Our group agreed that it can be used, in the following specific ways: As an immediate feedback system to individual students in conferences for self and professional development; for curricular improvement by immediate course revision; for long-range planning, or radically reshaping the program; for the political advantage of proving the program effective; and, for improvement of the Teacher Education profession by dissemination of standards, designs and procedures.

Knowing how the results can be used, the next logical question is, how can we incorporate program evaluation results into program components? There were several viable ways pointed out, including relating evaluation closely to course objectives and program goals. Furthermore, the objective-related evaluation results should be reported promptly to involved faculty, and assistance should be given to those faculty in interpreting results. We all agreed that the faculty should also be involved in the complete cycle of program development-implementation-evaluation-revision.

The third question is how to guarantee that results will be used. Our group decided that use of results must be planned for, making it a recognized and accepted program goal, and by cooperatively developing and formalizing procedures to attain that goal.

Those present identified as a critical prerequisite for the success of program evaluation a visible and acknowledged leader who takes responsibility for the full cycle of program evaluation activities. They also revealed a high level of interest in sharing knowledge, experience, instruments and procedures relating to program evaluation. There was an expressed desire for some assistance in determining what constitutes "ideal" use of program evaluation results. Perhaps some guidelines for "use" procedures could be generated and disseminated through TEPFU? Time precluded any further or more specific suggestions for the TEPFU group.

Participants

Darlene Huddleston
Tennessee State University

Charles R. Nash
Armstrong State College

Thomas H. Reeves
College of the Ozarks

Eugenia Stiemke
Valparaiso University

Waldo R. Widell
College of the Ozarks

PROGRAM EVALUATION--A SYSTEMS APPROACH

James J. Garland
Southwest Texas State University

Detroit, the Renaissance Center, and the Detroit Plaza were all grand, but they did not overshadow the period from 3:15 to 6:15 p.m. on February 18 in the Mackinac Ballroom. This session, which dealt with teacher education program evaluation, was very rewarding.

After hearing six very timely and informative papers, our discussion group convened for an exchange of ideas related to the presentations. We first took issue with the premise that the questionnaire technique of gathering information for program follow-up was not a good approach. However, we agreed that a systems approach was required for effective program study. Such a system would employ several techniques (including the questionnaire) for both short and longer periods of study. It was also agreed that personal interviews, public school evaluation teams, Teacher Centers, and students in the program serve as very effective components of the evaluation system.

Those in our group emphasized, as did the speaker, that evaluation of graduates is not, by itself, enough to support a strong evaluative process. The system should include the admissions office, departments, administrative offices, placement office, and other key components to provide a network of support agencies so that, each making its own contribution, a multi-approach system can be implemented.

The group agreed with the presenters when they pointed out that a sufficient funding base needs to be established for the evaluation system and that this base must be reflected in the regular budget process. Also, we supported the idea that a director or other individual should be designated with specific responsiblities of carrying out the policies and procedures for program evaluation and guiding and directing the activities.

The group supported the statement by Dr. Gubser that "...evaluation should not be an afterthought; it should be built into the program." We feel, as did the speaker, that evaluation must be planned as a part of the total teacher education program--not something added as an apparent need arises. The data gathered in the evaluation process serve as a guide for the program, as a sounding instrument for new and innovative ideas, as a measure of progress toward stated goals, and as a foundation for program accountability. Therefore, it is essential that the evaluation model be initiated as an integral part of the teacher education process.

As our group discussion progressed, we mentioned several ideas and component parts that would form a well planned and workable evaluation system. Due to the lack of time, however, our group did not formalize a workable model. On the basis of the six papers and our discussion of them, it would seem that Figure 1 represents a workable system for the agencies and their interrelationships in the evaluation process.

Participants

Walter Dukes
Mississippi Valley State University

Melvin F. Gadson
Winston-Salem State University

John Geiger
University of Dayton

Harry Hadley
Fairmont State College

Edell M. Hearn
Tennessee Technological University

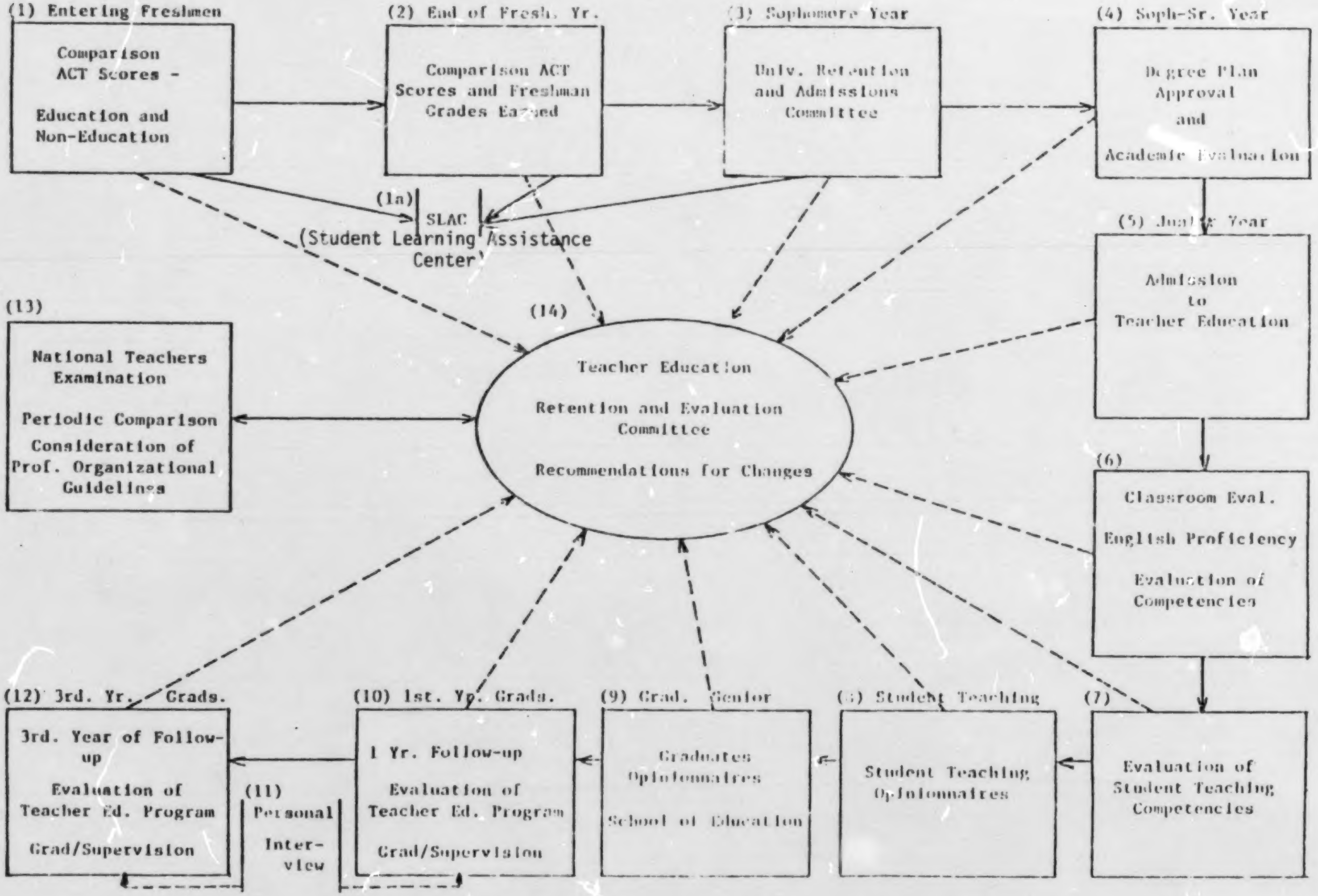
Judson P. Martin
Northeast Missouri State
University

James Murphy
Freed-Hardeman College

Gary Rush
University of Southern Mississippi

Edward B. Smith
Boston College

Figure 1
TEACHER EDUCATION EVALUATION
AND RETENTION PROCEDURES


(1) Entering Freshmen
Comparison ACT Scores - Education and Non-Education
(2) End of Fresh. Yr.
Comparison ACT Scores and Freshman Grades Earned
(3) Sophomore Year
Univ. Retention and Admissions Committee
(4) Soph-Sr. Year
Degree Plan Approval and Academic Evaluation
(1a) SLAC (Student Learning Assistance Center)
(5) Junior Year
Admission to Teacher Education
(13)
National Teachers Examination
Periodic Comparison
Consideration of Prof. Organizational Guidelines
(14)
Teacher Education Retention and Evaluation Committee
Recommendations for Changes
(6)
Classroom Eval.
English Proficiency
Evaluation of Competencies
(12) 3rd. Yr. Grads.
3rd. Year of Follow-up
Evaluation of Teacher Ed. Program
Grad/Supervision
(11) Personal Interview
(10) 1st. Yr. Grads.
1 Yr. Follow-up
Evaluation of Teacher Ed. Program
Grad/Supervision
(9) Grad. Senior
Graduates Opinionnaires
School of Education
(8) Student Teaching
Student Teaching Opinionnaires
(7)
Evaluation of Student Teaching Competencies

SMALL COLLEGE NEEDS IN FOLLOW-UP EVALUATION STUDIES

Al Kilgore
University of Nebraska-Lincoln

The follow-up small group session, facilitated by Al Kilgore, consisted of teacher education personnel from relatively small, private colleges. The primary focus, after introductions, centered on what each of the representative colleges were currently doing regarding follow-up studies, and more specifically, to describe the particular role each participant played in the evaluation process.

The level of involvement of the members of this group in follow-up studies of their graduates has been minimal, although all programs had both a semi-formal as well as informal network working for them. Informal and random visits to graduates who were teaching along with some observations of graduates in their classrooms, to questionnaires to first year teachers and interviewing graduates who had returned to campus for graduate classes were among the data gathering processes used. The data collected did not appear to be used systematically for program type decision making, rather the individuals collecting the data may have made some changes in the part of the programs that they had direct control over. For example, based on informal talks with graduates, one professor modified a methods course. At any rate, the final consensus was that each program had a follow-up evaluation process of some type and the data collected were helpful. Evaluation design, however, was very loose, unsystematic, and underfinanced.

This led to the issue that most concerned the participants in this group: to explore and learn evaluation designs that could be implemented (at a minimal cost) in their colleges. Participants were generally unfamiliar with the TEPFU publications and available information concerning the state of the art. Most would like to become part of a network and their names are included with this report.

The primary focus of the discussion involved how participants could get the "how to's" as to evaluation design and interpreting the results. Answers to the following questions and/or statements would be helpful to this group of participants.

+ Is there (should there be) some standardized data from follow-up studies so that one can compare a program to a standard?

+ Can "meta" analysis be performed on data already gathered? (Is there enough data?)

+ Are, or can there be, both norm referenced and criterion referenced data collected on follow-up studies? If so, what are some procedures?

+ Are there standards or statements of intent by states as to beginning teacher performance criteria? Should states set the standards? Can we get this information?

+ How do we develop "reference points" for developing evaluation designs and comparing graduates from different programs?

+ Should small colleges have evaluation standards, designs, and expectations different from large colleges?

+ Do current practices (such as described or inferred in the paper presentations) provide "good" data for decision making? Have programs actually changed as a result of such decision making? Is it really worth the effort?

+ Should outside evaluators be used to conduct follow-up studies, rather than program personnel who, in some cases, are the program?

+ Can NCATE standards be modified for programs of different size and emphasis? (This item was heavily discussed and there was general dissatisfaction with rule 6 of the NCATE standards as small colleges would have trouble meeting all criteria. Some colleges have already decided to pull out of the NCATE accreditation process because of these requirements.)

+ How does the "teaching" environment of the graduates influence follow-up studies? How can we account for this variable in our evaluations, assuming there is a difference?

+ Can long range (4-7 years) follow-up studies provide us with philosophical and general program data that can be useful for decision making and program modification? (The intent of the question appeared to center on the thought that the undergraduate program should have a longer-lasting effect on their graduates than one or two years, and we need to measure this effect.)

The general tone of the discussion and questions appeared to center on the need for participants to learn more of what is available, to share basic information, and to reinforce one another in the efforts to design sound follow-up studies and to use the data intelligently. Participants expressed the need for some "hands on" state or regional workshops that would help them answer the questions and develop acceptable evaluation designs.

The pragmatic needs expressed in this session included the need to impress their administrators to provide additional funds for evaluation purposes. This included the notions of training faculty as well as providing resources to conduct the follow-up studies (e.g., travel, computer time, released time, etc.). Perhaps TEPFU, vis-a-vis the Texas R&D Center, can develop an "evaluation program development kit" specifically designed for follow-up studies, and include cost estimates for various activities. This along with suggestions as to how to finance such projects appear to be needs as well as suggestions for further activities. This writer suggests that perhaps a whole teacher training program evaluation model(s) be developed which includes follow-up studies. It would appear that the state of the art is limited at this time.

Participants

Elmer Clauson
University of Pacific

Sister Rochelle Potaracke
Vertito College

Tammy Reid
Whitworth College

SO WHAT IS EVERYBODY TELLING US ABOUT EVALUATING TEACHER EDUCATION?

Joseph Vaughan
National Institute of Education
U. S. Department of Education*

Why Do We Need It?

Studying the six papers and eight small group reports from this AACTE session reminds me of the story of the blind man being placed next to an elephant and trying to figure out exactly what it was. His interpretation depended upon where he touched the beast and to which of its very different parts or dimensions he was exposed. So it appears to be with teacher education evaluation, whether one considers it a "beast" or a "beauty." We all have our own interpretations of what it should do and how it should do it. Some may look at teacher education evaluation and see it as principally dependent upon follow-up evaluations of graduates while others will view it as more of an internal review process which includes teacher education faculty and administrators and other advisors in a sort of introspective professional vigilance. Ask a dozen people and you'll probably get a dozen different interpretations, each with their own personal emphasis. Relating that to the story, it could be that we all may be at least partially "blind" to the total complexity and potential of evaluation in teacher education and thus find ourselves "feeling our way" and interpreting and emphasizing pieces rather than the whole entity. Hall suggests this in one place in his paper and I would like to expand that notion, examining the reasons for past and present weaknesses in order to derive suggestions of themes and processes for teacher evaluation. I will base these suggestions, for the most part, on information from the critiques and recommendations raised in the group of writings for this session and on my interpretations of the writings of others who have examined this topic. To a lesser extent, I shall be presumptuous enough to add my personal thoughts, identifying them as such. After all I'm fascinated with elephants, too!

---

*While the author is an employee of the National Institute of Education, opinions expressed herein are solely his own and do not necessarily represent the views of NIE or the U.S. Department of Education.

Perhaps the first reality we must all face is that the need for evaluation of teacher education is not going to go away. It is not only, as Gubser so succinctly articulates, a case of increased emphasis and guidance being put upon evaluation by NCATE. One does not have to be clairvoyant to predict where public dissatisfaction with education and teachers and the subsequent variety of state and locally initiated accountability measures will probably lead us. Just as local education agencies and the teachers within them are being publicly taken to task-- justifiably or unjustifiably--for perceived inadequacies, investigating some of the causes of those perceived inadequacies must include close examination of teacher education quality and quantity in process and content. The recent (April/May, 1981) series of Newsweek articles on the problems of our schools speak of teacher education as "generally, a hodgepodge of academic work, fieldwork and methodology which can be helpful tricks of the trade or irrelevant nonsense." The articles refer to a 56% failure rate among teacher applicants on the National Teacher Examination in one state, and in another, a 47% failure rate even though passing was set at about the 25th percentile. Weaver's work (1977) indicates that teacher education is attracting a less academically able population of students and much survey data exists to show us that teacher education institutions have done little to combat this trend or to establish anything more than minimal standards for admission and retention of teacher education students (AACTE, 1981). At the same time, a Newsweek poll reveals that 90% of the respondents felt that teachers should be required to pass a competency test of some type before they are hired. We have a wide gap which exists between public and, most of us would agree, logical professional expectations of at least minimal competence of teacher education graduates and the actual knowledge base which those graduates possess.

While the knowledge base which is assessed by most of the tests like NTE has not been shown to be correlated with effective teaching, logic mandates that we accept the fact that knowledge is a necessary, although not sufficient, condition of teaching competence. Clearly, teacher education graduates as a collective are not measuring up to what we as professionals must demand in the way of quality. We do not know the reasons why this is so and that is exactly why we need evaluation of teacher education efforts and graduates. We must understand how we can improve the present conditions to meet desirable and reasonable standards. Public dissatisfaction and professional obligation will only increase the importance of teacher evaluation in the future.

What Knowledge Do We Already Have and What Else Do We Need to Know?

The Adams/Craig paper on a national survey of teacher education evaluation practices offers many insights into what is presently taking place. Although some type of evaluation is used in almost 9 out of 10 institutions, the nature and comprehensiveness of these varies widely. Clearly, dominant practices include heavy reliance on questionnaires filled out by teacher education students, first year graduates, and to a somewhat lesser extent, supervisors. Much less use is being made of direct classroom observations or media-aided work sample observations, standardized tests as measures of knowledge or other factors, and feedback loops that tie evaluation back to program modification and improvement. Money supporting evaluation is minimal, with 50% of the institutions spending less than $1,250 annually for this purpose. Most evaluation is focused on the periods of preservice and first-year teaching follow-up. Many institutions had major concerns with "developing instrumentation," "identifying instrumentation," and "planning an evaluation system" in addition to the feedback problem mentioned earlier. Finally, "Teaching Skills, Strategies and Techniques," "Knowledge of Subject Matter" and "Relationships with Students" were areas of prime concern in the evaluation efforts.

This portrayal of the status quo is very enlightening. It reveals a reliance on superficial techniques that are minimal at best, a dearth of more comprehensive techniques (e.g., direct classroom observations), and very little financial commitment to evaluation. While teaching skills are identified as prime outcome concerns, very little use is made of evaluation measures that can yield valid and reliable data on these outcomes. For example, it is extremely difficult to imagine a questionnaire that would yield accurate and useful information on teaching skills, especially when the student or graduate her/himself is often the person being asked. Gubser also points out that return rates of 15% on such questionnaires are not uncommon. While we know from prior research on instruction that direct classroom observations and/or media-aided work samples are imperative, to understanding teaching and teacher effectiveness, these measures are glaringly absent from most teacher education student/graduate evaluations. One is led to believe, as was raised in one small group report, that cost considerations are the primary reasons for this absence, but more on that later.

Gubser makes an extremely significant point about the differences between and the need for both graduate and program evaluations and how the latter seems very much overlooked. He reminds us of the need to assess course sequences and content, textbook nature and use, admission and retention standards, and other program features as determinants of student acquisition of skills, knowledge, attitudes and behaviors while in preservice. He also discusses the importance of the roles and nature of involvement of teacher education faculty and administrators in planning, implementing, and using the results of evaluation activities. At the same time he warns us of the dangers of using graduate "success" or "competence" as surrogates for program "success" since we know very little about how much of the preservice effect "washes out" as a result of the unique influences of different school districts and schools in which graduates take positions. As Hall tells us, we must also be aware that the longer the graduate is in a regular teaching position, the more likely it will be that preservice effects are less potent. Certainly, graduate effectiveness has to be related to program effectiveness but they are discrete entitites which must also be understood separately before interrelationships can begin to be drawn with any degree of confidence. To date, there has been very little emphasis on these important distinctions.

Certainly, the concept of a teacher education continuum of which Hall speaks is as essential element in ultimately understanding the influence of preservice and inservice experiences and effectiveness. While each stage has its separate influences there is the ultimate need to determine how they are interrelated. This can be paired with recent emphasis on the need to examine this interrelatedness during the 1-3 year induction period of beginning teaching. Practitioners and researchers alike have identified induction as a period of high trauma and significant formative influence in how teachers will teach throughout their careers or, in many cases, whether or not they even want to stay in teaching. While there is no lack of quantity of data being collected on this period, especially the first year of teaching, the quality and comprehensiveness of the data are suspect.

Given these overview papers, the three papers by Ayers, Hearn and Craig/Adams offer useful, institution-specific cases which relate directly to the broader points of the other authors. The two institutions described are of particular interest because they are, in many respects, in the vanguard of those trying to systematically address teacher education evaluation. But they are also, I believe, in some respects plagued by the unsolved problems that are faced by the vast majority of teacher education institutions which are looking for help in designing and implementing evaluations.

Ayers and Hearn discuss efforts to develop a comprehensive model of graduate follow-up evaluations and have incorporated many desirable aspects mentioned in the three overview papers. The ten principles mentioned by Ayers reflect sound thinking about planning, implementation, use and revision issues in such evaluations. Key positive features include: 1) the need for both clarification of specific and general program objectives and relating evaluation to these; 2) a continuous and longitudinal design; 3) concern with advancements in evaluation methodology; 4) early and close collaboration with schools; 5) use of a variety of data collection techniques and sources and collection of both qualitative and quantitative data; and 6) a feedback procedure to make evaluation results known to faculty and administrators. Hearn adds the element of sufficient financial resource allocation to the activities and presents an excellent six-point analysis of key considerations including use of field-based faculty as data collectors, concern with the entire preservice/induction/inservice continuum, data sampling and processing issues, and multi-institutional consortia to pool resources and reduce costs.

While these are indeed positive and commendable features, they are in Ayers' words, "still incomplete and inadequate." (To some extent, one always faces the possibility that papers may simply not address some issues due to length restrictions rather than by conscious choice. I apologize in advance for citing concerns about these programs which may have been adequately addressed had the design of the papers been more amenable to telling the institutions' full stories.) For example, while the desirability of specifying objectives is unquestionable it is equally clear that much must also be known about the extent to which the teacher education practices matched the stated objectives. To move from objectives to evaluation without consideration of implementation produces a tenuous link at best. On the evaluation techniques issues, Ayers appears to confuse the progess made in techniques for research on effective teaching with techniques for effective evaluation. The former is very important as a determinant of the content of teacher education itself, but its linkage with evaluation processes is never specified. It is also important to consider, as far as school collaboration is concerned, how we can go beyond the "permission" and "understanding" levels to involving schools in the design, implementation and use of evaluation mechanisms which will be of as much benefit to school practitioners as to teacher educators. As far as program modification is concerned, the papers fail to address how the feedback loop is created and how the nature and extent of modifications is monitored. It is not sufficient to communicate results to faculty and administrators. There must be follow-up to see that something constructive happens through individual or group motivation as a result of the feedback. While Hearn's delineation of the importance and nature of evaluation costs is helpful, his suggestion that costs could be significantly reduced through elimination of observations is not appropriately qualified by the immense dangers of so doing. I would suggest that this element is at the heart of any potential contribution that could be made, and reliance on more superficial data collection is indefensible at this time. We simply do not have less expensive alternatives with which we can correlate teacher skills and performance. Finally, I would like to contest that principle number 9 in Ayers' paper may be the reverse of what should be the case. Specifically, is it not the case that program decision-making must provide for a workable system of evaluation and evaluation use rather than the other way around, as is stated? Evaluation is a part of program planning and implementation, not the reverse. It may be a subtle distinction but one of great importance in consideration of how evaluation can be fully integrated with program design, development, operation and modification.

The Craig/Adams paper shows many similarities to the Hearn and Ayers' descriptions, accurately reflecting the fact that there has been much collaboration and mutuality in the separate institution's development of evaluation efforts. In addition, Craig/Adams address the socio-political context within which teacher education evaluations exist. They also discuss the issues of use of such evaluations and the objective/subjective interpretation of results. A key concern is the hierarchical structure and conditions for making decisions about the nature and use of evaluations. The hierarchy emphasizes interaction and collaboration among teacher education faculty and administrators and advisory functions from school personnel. The ideal process for planning, implementing and using evaluations is seen as being based on "involvement" and "evolvement." Key final steps are checking the extent to which evaluations are used to make program modifications and determining, on a continuing basis, how the evaluation process can be improved.

Certainly the principles espoused in the Craig/Adams paper are laudable and much of the structure which I assume they have set up seems potentially appropriate to accomplishing their objectives. There do appear to be, however, alternative sequences of activities and combinations of decision-makers which might be considered in attempting to strengthen the evaluation program and, in particular, its impact. Initially, the assumption that one must start with the chief administrative officer (e.g., dean) to perform such functions as determining the "ideal program, need ..., options..., implications ..." could be questioned. Why not do this with a collaborative group, including the chief administrative officer, right from the beginning? In that fashion, a system of checks and balances would be in place, "ownership" would probably be enhanced, diverse perspectives would be ensured and none of the benefits of the chief administrator's sole determinations would be lost. What this really would amount to would be a reduction of three layers of decision-making into one with the chief administrative officer, decision-making team and planning/evaluation team memberships being represented on the composite team. The way the Craig/Adams hierarchy appears, there are two layers of decision-making before the "primary initiation and operational force" (planning/evaluation team) is brought into the picture. To separate establishment of "parameters" from "initiation" with two layers of decision bears the possibility of being perceived as a camouflaged "top-down" policy mandate. About the only reason I can perceive for not involving the role groups as collaborators from the beginning is the difficulty of large numbers, but it would appear that hard negotiation of a representative group would be small payment for ensuring removal of the potential "top-down" stigma.

Along with the above concern there are only two other reactions which jump to mind from the Craig/Adams paper. First, the delicate issue of the role of school practitioners has been addressed in the paper through having them in an "advisory" capacity. Given the realities of power structures in higher education, it is perhaps understandable that school personnel could not be given parity status in a formal sense. But classifying them as "advisory" could portend less than maximal involvement and perceptions that this input is more or less an "add-on." Any measures that could be utilized to ensure parity status, with or without formal position, should be encouraged. As most of us would agree, we are probably more likely to feel committed and contribute to an effort when we feel that our voice is of equal volume and impact as those with whom we are working. One other point deserves mention with regard to school personnel parity. It is those schools which are the "buyers" in a "buyers market." Does it not behoove teacher education personnel, and seem very reasonable, if only for survival, to pay full attention to the input of school personnel relative to 1) existing and changing contexts within which teacher education graduates will operate, and 2) the expectations and processes used by the "buyer" in determining which teacher education graduates are desirable and potential job-holders?

A second concern is with acceptance of the status quo "organizational framework" as the determinant of options. Realizing that one must acknowledge what exists and its limitations, that does not preclude working toward more desirable structures or frameworks. There is a "self-fulling prophecy" dimension to assuming you can't change simply because present structures won't allow it, just as there is in assuming no new sources of funds are likely for evaluation. One major difficulty identified and discussed by many (Smith, 1980; Denemark, 1979, etc.) is the nature of the present structure of teacher education. Without drastic changes in areas such as incentive structures, funding formulae from the state level, the college of education's control of its own governance, etc., we probably have less hope for exploration of major alternatives to present practice, including the area of evaluation and its impact. If a case is to be made for such changes (or at least the structural flexibility to consider them), we cannot afford a stance of passive acceptance of what exists. There is much worth keeping and much that needs examination and change. Our obligation would appear to be to work toward the structures within which such considerations and action are possible.

I must say, climbing down off my soap box, that the intent and substance of what Craig/Adams, Ayers and Hearn recommend are very much worth noting. My questions have to do with how everything can be most effectively and efficiently accomplished. Both institutions should be loudly applauded for their efforts to date and both serve as valuable guides to other institutions seeking possible answers to very difficult questions. My comments on their papers, as well as those of Gubser, Hall and the survey work of Craig and Adams, are intended to be as constructive as possible and with full recognition of and respect for the considerable work and thoughts evident in each paper.

What Are Some "Best Guesses" At This Time About Where We're Headed in Teacher Education Evaluation Efforts?

I would like to talk a bit about what we are likely to see happening in the near and distant future and how that might compare with what some might consider "ideal" conditions. In order to do this, it appears to me as well as to many of the writers, that we must consider not only follow-up studies but the whole spectrum of teacher education program and product assessments from the beginning of undergraduate work through some specified period after graduates are teaching in the schools. Without this conceptualization of concern for the total effort over a continuum of time, the basic points of Hall, Gubser and several others will go unaddressed.

In considering the future, I will try to incorporate the responses of the eight small group discussions along with my own thoughts. I would be very remiss if I did not comment on how impressive the brief small group reports were, given the short time they had to work and the complexity of the issues. While the reports indicate that the topic itself is one of great currency, interest and controversy, they also indicate the considerable potential of this type of interaction in producing worthwhile thinking and suggestions. The diversity and excellence of thought is heartening.

Several of the comments already made have obvious implications for the future. It is not necessary to reiterate those in-depth, only to summarize that the core of present and future teacher education evaluation efforts must be formed by more comprehensive, detailed and objective data collection; collaboration on design, implementation and use; concern with both program and products; adequate financial support; and a longitudinal perspective. It is also clear that this is frequently not the case in existing efforts. The small group reports give us a broader perspective on the reasons why, as well as what is needed to begin moving toward improvement of the quality and quantity of evaluations. I shall refer to the reports by the name of the discussion leader in the group.

Virtually all of the group reports substantiated many of the basic weaknesses and strengths brought out by presenters. But each group took its own approach in saying what that meant to them and how they saw themselves addressing the problems and capitalizing upon existing "success" and knowledge. Perhaps the central concern of the groups was best stated by the Dravland comment that what was needed was "a well articulated framework within which we can define our tasks and our areas of ignorance."

The White report emphasized the need for initially considering the match or mismatch between "professional community needs" and "program outcomes" and, within that, the critical point of identifying instructional and program objectives. One difficulty with that, as the Froyen group discusses, is that the faculty who comprise much of that community are highly resistant to change, often in the name of academic freedom and autonomy. Froyen contends that little attention is paid to consensus-building in developing objectives and even less attention to standards of compliance with those objectives. So we have a built-in conflict in establishing a sound foundation for developing reasonable and appropriate evaluation questions and processes. In the past this has often resulted in no action. Fortunately, (or unfortunately, depending on your perspective), there are now accountability forces in effect that may necessitate action. Froyen's excellent list of reasons commonly given about why we can't do quality evaluations are likely to be overwhelmed by the state legislative and state department of education mandates, professional association activities and accrediting agency demands for "proof" that teacher education is doing its job satisfactorily. In short, more stringent measures of "success" will be mandated and, put most simply, it is logical to assume that university faculty and the rest of the professional community would rather get involved and have some say in their own destiny as opposed to having their fate totally determined by other external forces. This is perhaps not the most positive or altruistic reason for becoming involved, but it is likely to be a reality. Once we move beyond specification and clarification of objectives, we find ourselves in a position of having to know what presently exists in teacher education evaluation practices. A common information base is a necessity if we are to avoid "reinventing the wheel," as the deVoss group discusses, and capitalize upon past successes and failures. The kinds of questions that must be asked and for which information must be gathered are typified in the Kilgore report. What are present standards for graduates which are being mandated by various states to which our students might go for a job? How does NCATE fit in with all this? Have some teacher education programs actually changed as a result of evaluations and, if so, how? How do we look at the impact of preservice given the uncertain effects of each unique school setting? These and other questions posed by Kilgore are somewhat scattered but that is exactly the real world for most teacher education institutions. Institutions have been too isolated and need to become aware of the state-of-the-art of evaluation and the context within which that evaluation must exist. As White states, we need descriptions of "evaluation programs currently in place or under development with a discussion of conceptual features, operational preconditions, implementation problems and benefits, cost effectiveness and, most importantly, what difference has been made as a result of the program."

In terms of the actual design and conduct of evaluations, many good suggestions were made by the groups. For example, the deVoss paper contains several creative ideas including the use of volunteer teachers sitting on-campus with faculty groups to discuss evaluation issues, documentation beginning early in undergraduate years, "critical incident" logs being kept by teacher education students, cooperating teachers and supervisors, and intense data collection with smaller random samplings of students/graduates. NCATE itself, deVoss suggests, has been a somewhat unintentionally negative influence due to its demands for what many institutions see as a limiting view of evaluation. At the same time, it has the potential to be a very positive influence if it could play a more active role in disseminating information about successful approaches and encourage experimentation and diversification in evaluating programs and products. Certainly, the Gubser presentation indicates very harmonious thinking along those lines and, perhaps portends just such activity. Again, a basic issue to be addressed is exactly just what constitutes "proof" that teacher education is doing its job and to understand teacher education's limitations as well as its potential. The Savage group (sounds menacing, doesn't it?) reminds us that the systems model used in industry guarantees the "product" only as long as it is used as intended. The potential mismatch between "ideal" preparation and the idiosyncratic realities of individual schools mandates that institutions of teacher education closely examine their objectives and programs within that real world context. The "proof" indicators must emanate from a realistic assessment of the relevance and impact of preservice on what a teacher does day-to-day in a classroom, school and community environment. The idea of more active state involvement in teacher evaluation, as in the Georgia example cited in the Savage paper, must be investigated and understood in relationship to the evaluative roles of the higher education institutions and the schools.

Finally, in terms of the use of evaluation results Hungerman raises a very important point about the need to formalize processes for communication, discussion and modification of program if we are to really bring about changes. To just disseminate the results to faculty and administrators is not enough. It is extremely important, the report goes on, to have a visible and acknowledged (and, I would add, respected) leader for the evaluation effort to maximize impact. It strikes me that one alternative to having to choose between an evaluation or program "type" for this role is to consider dual leadership where both roles would be represented. Through this means, credibility is likely to be increased along with the overall relevance and validity of the effort. Garland reminds us, as did others, that whatever results are used, they must reflect evaluation of the total teacher education effort from admissions to program replacement and beyond. Only in that way can we consider evaluation truly systematic.

In summary, we have our work cut out for us! We've identified a great number of serious weaknesses but we've also identified many useful approaches and principles worthy of consideration. The time has come to move beyond stating the problems and to look at how we can build upon the conceptual and structural necessities and beginnings discussed with much insight in these papers. We need not be too cynical or pessimistic. This session reveals a valuable pool of knowledge and judgment about what is needed and how we can begin to move toward those ends. It may be a difficult task we are undertaking but it need not be overwhelming. Awareness has been created and we now must move toward action. This work stands as an excellent beginning.

References

American Association of Colleges for Teacher Education. AACTE Briefs, January-February, 1981, 2(1), 5-7.

Denemark, G. The case for extended programs of initial teacher preparation. Presentation to the Forum of Education Organization Leaders, Arlington, Virginia, September, 1979.

Smith. B. O. A design for a school of pedagogy. U.S. Department of Education, U.S. Government Printing Office, Washington, DC. Publication No. E-80-42000, 1980.

Weaver, W. T. Educators in supply and demand: Effects on quality. Presentation to American Educational Research Association, New York, April, 1977. (ERIC No. ED 143 145).